Density Estimates for a Random Noise Propagating through a Chain of Differential Equations

François Delarue, Stéphane Menozzi

▶ To cite this version:

François Delarue, Stéphane Menozzi. Density Estimates for a Random Noise Propagating through a Chain of Differential Equations. Journal of Functional Analysis, 2010, 259, pp.1577–1630. hal-00436051

HAL Id: hal-00436051
https://hal.science/hal-00436051

Submitted on 25 Nov 2009

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Density Estimates for a Random Noise Propagating through a Chain of Differential Equations

François Delarue [a], Stéphane Menozzi [b],

[a]*Laboratoire J.-A. Dieudonné, Université de Nice Sophia-Antipolis, Parc Valrose, 06108 Cedex 02, Nice, France*

[b]*Laboratoire de Probabilités et Modèles Aléatoires, Université Paris 7-Diderot, Case 7012, Site Chevaleret, 5 rue Thomas Mann, 75205 Paris Cedex 13, France*

Abstract

We here provide two sided bounds for the density of the solution of a system of n differential equations of dimension d, the first one being forced by a non-degenerate random noise and the $n-1$ other ones being degenerate. The system formed by the n equations satisfies a suitable Hörmander condition: the second equation feels the noise plugged into the first equation, the third equation feels the noise transmitted from the first to the second equation and so on..., so that the noise propagates one way through the system.

When the coefficients of the system are Lipschitz continuous, we show that the density of the solution satisfies Gaussian bounds with non-diffusive time scales. The proof relies on the interpretation of the density of the solution as the value function of some optimal stochastic control problem.

Key words: Aronson estimates, Gaussian bounds, Hypoellipticity, Hörmander Conditions, Stochastic Control.

Email addresses: `delarue@unice.fr` (François Delarue), `menozzi@math.jussieu.fr` (Stéphane Menozzi).

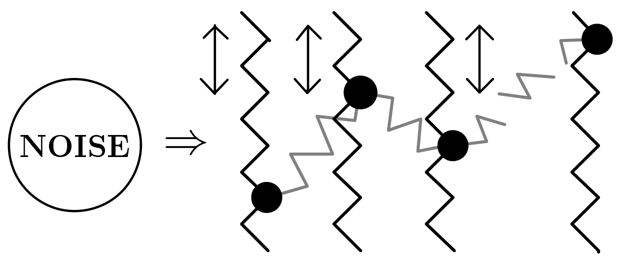
URLs: `http://math.unice.fr/~delarue` (François Delarue), `http://www.proba.jussieu.fr/~menozzi` (Stéphane Menozzi).

1 Introduction

We are here interested in understanding how a random noise may propagate through a chain of n d-dimensional differential equations of the form

$$
\begin{aligned}
dX_t^1 &= F_1(t, X_t^1, \dots, X_t^n)dt + \sigma(t, X_t^1, \dots, X_t^n)dW_t, \\
dX_t^2 &= F_2(t, X_t^1, \dots, X_t^n)dt, \\
dX_t^3 &= F_3(t, X_t^2, \dots, X_t^n)dt, \qquad\qquad t \geq 0, \qquad (1.1) \\
&\dots \\
dX_t^n &= F_n(t, X_t^{n-1}, X_t^n)dt,
\end{aligned}
$$

$(W_t)_{t\geq 0}$ standing for a d-dimensional Brownian motion, and each $(X_t^i)_{t\geq 0}$, $1 \leq i \leq n$, being $\mathbb{R}^d$-valued as well. A typical example for (1.1) is a system of n coupled oscillators, each of them moving vertically and being connected to the nearest neighbours directly, the first oscillator being forced by a random noise. Such a model may be summarized by the picture



We emphasize that F_i in (1.1) cannot depend on the positions $X_t^1, \dots, X_t^{i-2}$: from a physical point of view, the noise has to go through the second, the third, ... and the $(i-1)^{\text{th}}$ oscillators to reach the i^{th} one. For example, the picture given above corresponds to the typical case when $F_1 = F_1(t, x_1, x_2)$, $F_2 = F_2(t, x_1, x_2, x_3)$, $F_3 = F_3(t, x_2, x_3, x_4)$ and so on... This kind of systems appear in various applicative fields. When $n = 2$, Eq. (1.1) describes the dynamics of some stochastic Hamiltonian systems (see e.g. Soize [Soi94] for a general overview or the more specific works by Talay [Tal02] and Hérau and Nier [HN04] for questions of convergence to equilibrium). Again for $n = 2$, it corresponds to the dynamics used in mathematical finance to price Asian options (see for example [BPV01] for a specific discussion of the regularity of the price in such a degenerate case). In the more general case when $n \geq 2$, it appears in heat conduction models (see for example the original papers by Eckmann et al. [EPRB99] and Rey-Bellet and Thomas [RBL00] when the chain is forced by two heat baths ; see also the more recent paper by Bodineau and Lefevere [BL08]).

Hörmander Setting. When the coefficients of (1.1) are smooth, the existence of a density for $(X_t^1, \dots, X_t^n)$, seen as an $\mathbb{R}^{nd}$-valued vector, may be seen as a consequence of Hörmander's theorem. For simplicity reasons, we just explain how when $d = 1$, i.e. when each oscillator is of dimension 1. In the one-dimensional case, the Dynkin operator associated with (1.1) has the form (below $F_1(t, x_0, \cdots, x_n) = F_1(t, x_1, \cdots, x_n)$)

$$\partial_t + \mathcal{L}_t = \partial_t + \frac{1}{2}\sigma^2(t, x_1, \dots, x_n)\partial^2_{x_1,x_1} + \sum_{i=1}^n F_i(t, x_{i-1}, \dots, x_n)\partial_{x_i} = \frac{1}{2}A^2 + B, \tag{1.2}$$

with $A := \sigma(t, x_1, \dots, x_n)\partial_{x_1}$ and $B := \partial_t + (F_1 - \frac{1}{2}\sigma\partial_{x_1}\sigma)(t, x_1, \dots, x_n)\partial_{x_1} + \sum_{i=2}^n F_i(t, x_{i-1}, \dots, x_n)\partial_{x_i}$ for $i \geq 2$. Then, the Lie algebra of the vector fields A and B contains $B = (1, \times, \cdots, \times)$ (the first coordinate stands for time t), $A = (0, \sigma, 0, \dots, 0)$, $[A, B] = (0, \times, \sigma\partial_{x_1}F_2, 0, \dots, 0)$ as well as

$$\begin{aligned}
&[[A, B], B] = (0, \times, \times, \sigma\partial_{x_1}F_2\partial_{x_2}F_3, 0, \dots, 0),\\
&\dots\\
&\Big[\cdots\big[[A, B], B\big], \underbrace{\cdots, B}_{(n-3)\text{ times}}\Big] = (0, \times, \cdots, \times, \sigma\partial_{x_1}F_2\,\partial_{x_2}F_3 \cdots \partial_{x_{n-1}}F_n).
\end{aligned}$$

Above, each cross "$\times$" means that the corresponding value doesn't provide any useful information to span new directions. If, at any $(t, \mathbf{x} = (x_1, \dots, x_n)) \in (0, +\infty) \times \mathbb{R}^n$, the terms $\sigma(t, \mathbf{x})$, $\partial_{x_1}F_2(t, \mathbf{x})$, $\cdots$, $\partial_{x_{n-1}}F_n(t, x_{n-1}, x_n)$ are non-zero, then the vector fields B, A, $[A, B]$, $[[A, B], B]$, $\cdots$, $[\cdots[[A, B], B]\cdots, B]$ span $\mathbb{R}^{n+1}$ at any $(t, \mathbf{x})$. By Hörmander's theorem, this implies that the operator $\partial_t + \mathcal{L}_t$ is hypoelliptic on $(0, +\infty) \times \mathbb{R}^n$. By the same argument, the adjoint operator is hypoelliptic as well. This implies that, for any starting point $(t, \mathbf{x})$ as above, the process $(X_s^1, \dots, X_s^n)_{s \geq t}$ has a density at any time $s > t$: despite the degeneracy of the $n - 1$ last oscillators, the whole state of the system is (strictly) random at any time after t because of the propagation of the noise through the oscillators. Such a principle may be extended to the case when the oscillators are d-dimensional. Indeed, if, at any $(t, \mathbf{x} = (x_1, \dots, x_n)) \in (0, +\infty) \times (\mathbb{R}^d)^n$, the $\mathcal{M}_d(\mathbb{R})$-matrices [1] $\sigma(t, \mathbf{x})$, $D_{x_1}F_1(t, \mathbf{x})$, $D_{x_{n-1}}F_n(t, x_{n-1}, x_n)$ are of non-zero determinant, then Hörmander's theorem still applies.

Gaussian Case. Providing two-sided bounds for the density, when it exists, of the random vector $(X_t^1, \dots, X_t^n)$, $t > 0$, is a natural question, which goes back to the earlier work of Kolmogorov [Kol34]. Indeed, for $d = 1$, $n = 2$, $\sigma = 1$, $F_1 = 0$ and $F_2(x_1, x_2) = \alpha x_1$, with $\alpha \neq 0$, $(X_t^1, X_t^2)_{t \geq 0}$, with $(X_0^1, X_0^2) = (x_1^0, x_2^0)$, has the form $(x_1^0 + B_t, x_2^0 + \alpha x_1^0 t + \alpha \int_0^t B_s ds)_{t \geq 0}$: this example, known as Kolmogorov's example, is the first historical illustration of the noise propagation property for equations of type (1.1). Indeed, $(X_t^1, X_t^2)_{t \geq 0}$

[1] $\mathcal{M}_d(\mathbb{R})$ is the set of square matrices of size d.

is a Gaussian process whose covariance matrix at time $t > 0$ is equal to

$$K_t = \begin{pmatrix} t & \alpha t^2/2 \\ \alpha t^2/2 & \alpha^2 t^3/3 \end{pmatrix}$$

and is thus non-degenerate. Clearly, (X_t^1, X_t^2) has a density for any $t > 0$, which is explicitly given by

$$(x_1, x_2) \mapsto \frac{\sqrt{3}}{\alpha \pi t^2} \exp\Big(-\frac{|K_t^{-1/2}(x_1 - x_1^0, x_2 - x_2^0 - t\alpha x_1^0)^*|^2}{2}\Big). \tag{1.3}$$

We emphasize that the time-scale in the density is thus non-diffusive: the exponent of the fluctuations of the second component is $3/2$. As explained in Section 3 below, Kolmogorov's example may be extended to systems of n linear oscillators of dimension d. The typical case corresponds to the following choice for the coefficients $F_2, \dots, F_n$: $F_1 = 0$, $F_2(t, x_1, \dots, x_n) = \alpha_t^1 x_1$, $F_3(t, x_2, \dots, x_n) = \alpha_t^2 x_2$, $\dots$ and $F_n(t, x_{n-1}, x_n) = \alpha_t^{n-1} x_{n-1}$, $\alpha_t^1, \dots, \alpha_t^{n-1}$ standing for matrices in $\mathcal{M}_d(\mathbb{R})$. In this situation, the process $(X_t^1, \dots, X_t^n)_{t \geq 0}$ is Gaussian. When the matrices $\alpha_t^1, \dots, \alpha_t^{n-1}$ are non-degenerate, the covariance matrix of $(X_t^1, \dots, X_t^n)_{t \geq 0}$ has a non-zero determinant and, for each $1 \leq i \leq n$, the fluctuations of the i^{th} component is exactly of order $t^{(2i-1)/2}$.

Aronson Estimate. Of course, the linear case is very restrictive for practical purposes. Unfortunately, understanding the nonlinear setting is a much more challenging question. When the system is made of a single oscillator, the problem just consists of estimating the density of a non-degenerate diffusion process, the matrix σ being indeed assumed to be non-degenerate as explained in the previous paragraph. This problem has been widely investigated in the literature for fifty years. When F_1 (which just depends on (t, x_1) in this case) and σ are bounded and Hölder continuous in space, $\sigma\sigma^*$ being also uniformly elliptic, it can be derived from Friedman [Fri64] that the process $(X_t^1)_{t \geq 0}$ admits a transition kernel $(p(t, x_1, y_1))_{t>0, x_1 \in \mathbb{R}^d, y_1 \in \mathbb{R}^d}$, which satisfies on any interval $(0, T]$, $T > 0$, two-sided Gaussian bounds of the form $C^{-1} t^{-d/2} \exp(-C|x_1 - y_1|^2/t) \leq p(t, x_1, y_1) \leq C t^{-d/2} \exp(-C^{-1}|x_1 - y_1|^2/t)$, known as Aronson estimates [2]. Here, C only depends on the bounds and the regularity of the coefficients (including the lower bound of the spectrum of the diffusion matrix), on d and on T. The upper bound as well as controls on the derivatives of the heat kernel are given in [Fri64], the lower bound then follows from a standard chaining argument, see also Section C. The proof of the upper bound is based on a parametrix representation of the density deduced from a perturbation argument. We refer to [Fri64], [MS67] or Section 2.4 for

[2] Actually, Aronson's work [Aro67] deals with divergence form operators. In particular, it is shown that the transition densities of a self-adjoint non-degenerate operator satisfy Gaussian bounds, uniformly in time.

additional details on parametrix techniques. The two-sided bounds have also been derived in this framework by Sheu [She91] using stochastic control tools.

Density Estimate in the Hypoelliptic Setting. Understanding the structure of the density under general hypoelliptic conditions (i.e. for more general systems than the one we here analyze) is something very difficult. The reason may be explained in a simple way: there may be many ways for the underlying noise to propagate into the whole system, and therefore, many different time scales for the propagation phenomenon [3]. Nevertheless, several results have been obtained by means of Malliavin calculus: Malliavin calculus (see Malliavin [Mal78b, Mal78a, Mal97], Stroock [Str83] and Nualart [Nua95]) permits to quantify the sensitivity of the system with respect to the noise, the sensitivity being read through the Malliavin derivatives and summarized by the so-called "Malliavin matrix". Asking for the existence of a density is asking for a non-zero sensibility and thus for a non-degenerate "Malliavin matrix". The most famous work in that direction is due to Kusuoka and Stroock in a series of three papers [KS84, KS85, KS87]: in the last one, explicit two-sided bounds for the density of time-homogeneous diffusion processes are established by Malliavin calculus under strong Hörmander conditions. Roughly speaking, "strong Hörmander" means that the noise propagates inside the system through the diffusive part only: the density estimates established in Kusuoka and Stroock [KS87] hold for a null drift only [4]. On the opposite, the current problem we here consider is of weak Hörmander type since the drift has a key role in the noise propagation.

For other examples of application of Malliavin calculus, we refer the reader to Bally [Bal90], where the connection between the "Malliavin matrix" and the Hörmander assumption is investigated carefully, to Cattiaux [Cat90], where various properties of the resolvent of a diffusion process are established in the Hörmander setting, and to Cattiaux and Mesnager [CM02] for a careful analysis of the non-homogeneous framework. For a more specific result on the shape of the density, we refer to Ben Arous and Léandre [BL91], where various regimes are exhibited for the small time behaviour of the density of some hypoelliptic diffusion according to the form of the drift.

Our situation may seem more favorable: as already said, the noise propagation is one-way ; moreover, as explained below, the noise is transmitted from one oscillator to another in a regular (or non-singular) way, i.e. the Jacobian matrix $D_{x_{i-1}}F_i(t, x_{i-1}, x_i, \ldots, x_n)$ is assumed to be non-degenerate, uniformly in space and time. In this framework, several results have been already obtained. For

[3] On the opposite, when the system just consists of a single uniformly elliptic operator, there is one rate of propagation given by the exponent $1/2$.

[4] Actually, a very specific extension is discussed in [KS87] in the case when the drift is generated by the vector fields of the diffusive part: we refer the reader to the original paper for more details.

example, when F is linear, upper bounds were established in the ealier works by Weber [Web51] for $n = 2$ and by Sonin [Son67] for $n = 3$. Moreover, in the linear case again (but for a general n), lower bounds were obtained by Pascucci and Polidoro [PP06] and by Boscain and Polidoro [BP07] by using techniques involving Harnack inequalities and deterministic optimal control. Eventually, for a nonlinear F, Bally and Kohatsu-Higa [BKH09] have recently obtained lower bounds for $n = 2$ and $d = 1$ by means of Malliavin calculus. We notice that, as far as the lower bound is concerned, all the above results hold for time-homogeneous coefficients only.

Our approach. To derive two sided Gaussian bounds in the framework of equation (1.1), we follow a stochastic control approach. It consists in giving a representation of $-\ln p(T, \mathbf{x}, \mathbf{y})$, where p stands for the density at point $\mathbf{y}$ of $\mathbf{X}_T$ starting from $\mathbf{x}$ at time 0, as the value function of a stochastic control problem. Such a procedure, known as Fleming's transform has been successfully used by Sheu [She91] to derive Aronson's bounds as well as bounds on the derivatives of the heat kernel in the non-degenerate case.

In our degenerate framework we derive the lower bound for $p(T, \mathbf{x}, \mathbf{y})$ by taking some specific control, whereas the proof of the upper bound relies on both parametrix and stochastic control techniques. Namely, the control formulation is used to truncate the parametrix series expansion of the density. The strategy we use to derive our main results is specified in Section 2. For completeness, we also provide in Appendix an alternative proof of the lower bound on the same model as the proof of the upper bound: at first sight, it seems much shorter than the main (or the first) proof, given in the core of the paper. Actually, the following has to be said: first, we feel interesting to have two different strategies at hand for the lower bound since lower bounds are usually known as difficult to obtain ; second, some of the arguments of the second proof follow from the first proof, namely those related to the stochastic control approach, some others also follow from the proof of the upper bound, namely those related to the parametrix method ; this may explain why the second proof seems to be so short ; third, and this is probably the most important point, we think that the first proof is more systematic than the second one: the whole problem is to seek for a specific control to bound from below the density ; in other words, the first proof might be used in other situations more easily.

Standing Assumption. The required assumption in our framework is

(A) The spectrum of the matricial function $(t, \mathbf{x}) \in \mathbb{R}_+ \times \mathbb{R}^{nd} \mapsto a(t, \mathbf{x}) = [\sigma\sigma^*](t, \mathbf{x})$ is included in $[\Lambda^{-1}, \Lambda]$ for some $\Lambda \geq 1$. Moreover, the functions $F_1, \dots, F_n$ and σ are respectively uniformly Lipschitz and η-Hölder continuous ($\eta \in (0, 1]$) with respect to the underlying space variables, for some positive constant κ. For each integer $2 \leq i \leq n$, $(t, (x_i, \dots, x_n)) \in \mathbb{R}_+ \times \mathbb{R}^{(n-i+1)d}$, the function $x_{i-1} \in \mathbb{R}^d \mapsto F_i(t, x_{i-1}, \dots, x_n)$ is also continuously differentiable, the derivative, denoted by $(t, x_{i-1}, \ldots, x_n) \in \mathbb{R}_+ \times \mathbb{R}^{(n-i+2)d} \mapsto D_{x_{i-1}} F_i(t, x_{i-1}, \ldots, x_n)$, being η-Hölder continuous in the first space variable x_{i-1} with constant κ. Finally, there exists a closed convex subset $\mathcal{E}_{i-1} \subset GL_d(\mathbb{R})$ (set of invertible $d \times d$ matrices) s.t., for all $t \geq 0$ and $(x_{i-1}, \ldots, x_n) \in \mathbb{R}^{(n-i+2)d}$, the matrix $D_{x_{i-1}} F_i(t, x_{i-1}, \ldots, x_n)$ belongs to $\mathcal{E}_{i-1}$. For example, $\mathcal{E}_i$, $1 \leq i \leq n-1$, may be a closed ball included in $GL_d(\mathbb{R})$, which is an open set.

We notice that the coefficients may be irregular in time ($F(t, \mathbf{0})$ bounded). Obviously, they are assumed to be measurable. Also, the last part of Assumption (**A**) will be explained in Section 3.

Weak Solvability. Before we provide two-sided bounds for the transition density of the solution of (1.1), we emphasize that the unique solvability of (1.1) under Assumption (**A**) is actually not so obvious. The point is the following: if (1.1) is known to be uniquely solvable in the weak sense, we can prove that the transition densities exist and satisfy Gaussian bounds depending on the parameters in (**A**) only. Specifically, existence of a weak solution is easily seen: it can be proven by compactness argument as in Theorem 6.1.7 in Stroock and Varadhan [SV79]. (Note that the linear growth of the drift is not a problem to do so.) Uniqueness is here necessary also since the bounds for the density are obtained first for smooth coefficients and then deduced in the general case by a convergence in law argument. Actually, we do not know whether the martingale problem associated with (1.1) is indeed uniquely solvable under (**A**). We feel that it is, but we have no rigorous argument for it. The proof would require a careful analysis which seems out of scope here. We thus prefer to postpone the study of uniqueness under (**A**) to further investigations and to provide several interesting examples for which uniqueness holds:

Example 1. Obviously, if the coefficient a is locally Lipschitz continuous in space, uniformly in time, pathwise uniqueness holds.

Example 2. A possible way to relax the Lipschitz continuity property for a consists in taking advantage of the theory of viscosity solutions (for partial differential equations). Indeed, as recently shown by Ma and Zhang [MZ08] in a paper devoted to weak uniqueness of the Feynman-Kac representation of Backward SDE type for some non-linear partial differential equations, weak uniqueness may be seen as a consequence of the comparison principle for viscosity solutions. Unfortunately, it is inopportune to reproduce the whole theory in our (long) paper. Nevertheless, we point out that the argument given by Ma and Zhang can be adapted to the linear framework easily: as suggested by comparing Theorem 5.5 in [MZ08] with Theorem 6.2.3 in Stroock and Varadhan [SV79], it is well understood that the unique solvability of the martingale problem follows from the comparison principle for viscosity sub– and supersolutions of the linear PDEs driven by $(\mathcal{L}_t)_{t\geq 0}$. (Specifically, we should say for the Cauchy problems driven by $(\mathcal{L}_t)_{t\geq 0}$ with a smooth boundary condition and a null source term, the expression "Cauchy problems" being possibly understood as "Cauchy problems on bounded domains": in light of Theorem 6.6.1 in Stroock and Varadhan [SV79], weak uniqueness can be localized easily.) A common reference for comparison principles for viscosity solutions is the paper by Ishii and Lions [IL90]: putting together Theorem III.1 and Paragraphs IV.1 and IV.2 therein, we understand that, for a, F_1, …, F_n continuous both in t and x, the comparison principle holds for $\eta \in (1/2, 1]$. In other words, we claim that the martingale problem is well-posed when $\eta \in (1/2, 1]$ in (**A**) and the coefficients are continuous in time.

Example 3. As said above, we don't know whether the case $\eta \in (0, 1/2]$ is reachable. We feel that the right idea would consist in adapting the Schauder theory (see [Fri64] for the original theory in the uniformly elliptic setting) to establish existence of a classical solution to the Cauchy problem driven by $(\mathcal{L}_t)_{t\geq 0}$ (and by a smooth boundary condition): as already said, such a program is out of the scope of the paper. Nevertheless, we know that it can be achieved in the case when each F_i, $2 \leq i \leq n$, is independent of t and linear in $(x_{i-1}, \dots, x_n)$: we refer to the papers by DiFrancesco and Pasucci and DiFrancesco and Polidoro [FP05] and [FP06]. (Specifically, we refer to Theorem 1.4 in the first reference and Corollary 1.4 in the second one.) In that case, the Cauchy problem (set on the whole space) driven by $(\mathcal{L}_t)_{t\geq 0}$ and by a continuous boundary condition is known to be solvable in the strong sense whenever the coefficients a and F_1 are bounded and (t, x)-Hölder continuous for the "intrinsic geometry induced by the vector fields" (see the definition in the previous references): following Theorem 6.3.2 in Stroock and Varadhan [SV79], weak uniqueness follows. By the Girsanov transform, it is enough to prove uniqueness of the martingale problem when F_1 is zero, so that F_1 may be assumed to be Hölder continuous as required. As far as a is concerned, it can be checked that "usual Hölder continuity and boundedness" implies "intrinsic Hölder continuity" (see Proposition 2.1 in Polidoro [Pol94]). Finally, we claim that weak uniqueness holds when each F_i, $2 \leq i \leq n$, is linear and a is (t, x)-Hölder continuous, the Hölder exponent being possibly less than 1/2.

Main Result. We now state the main result of the paper:

Theorem 1.1 *Assume that Assumption (**A**) is in force and that weak uniqueness holds for Eq.* (1.1) *for any initial condition. (See above for possible examples.) Then, at any time $t > 0$ and for any initial condition $\mathbf{x} \in \mathbb{R}^{nd}$, the law of the vector $(X_t^1, \dots, X_t^n)$, solution at time $t > 0$ of the equation* (1.1) *under the initial condition $(X_0^1, \dots, X_0^n) = \mathbf{x}$, admits a density $\mathbf{y} \in \mathbb{R}^{nd} \mapsto p(t, \mathbf{x}, \mathbf{y})$. Moreover, for any $T > 0$, there exists a constant $C_T \geq 1$, depending on T, Λ,*

η, κ, $\mathcal{E}_1$, $\mathcal{E}_2$, ..., $\mathcal{E}_{n-1}$, n and d only, such that, for any $t \in (0,T]$,

$$\begin{aligned}
&C_T^{-1} t^{-n^2 d/2} \exp\Big(-C_T t \Big|\mathbb{T}_t^{-1}\Big(\boldsymbol{\theta}_t(\mathbf{x}) - \mathbf{y}\Big)\Big|^2\Big) \\
&\quad \le p(t, \mathbf{x}, \mathbf{y}) \\
&\quad \le C_T t^{-n^2 d/2} \exp\Big(-C_T^{-1} t \Big|\mathbb{T}_t^{-1}\Big(\boldsymbol{\theta}_t(\mathbf{x}) - \mathbf{y}\Big)\Big|^2\Big).
\end{aligned}$$

Above, $\mathbb{T}_t$ is the "scale" matrix of the system. It is a diagonal matrix of size $nd \times nd$ given by n diagonal blocks of size $d \times d$, with $t^i I_d$ as i^{th} diagonal block, where I_d is the identity matrix of size $d \times d$. Moreover, $\boldsymbol{\theta}_t(\mathbf{x}) = (\theta_t^1(\mathbf{x}), \theta_t^2(\mathbf{x}), \dots, \theta_t^n(\mathbf{x}))$ stands for the value (in $\mathbb{R}^{nd}$) at time t of the solution of the deterministic ODE

$$\begin{aligned}
\dot{\theta}_t^{\,1} &= F_1(t, \theta_t^1, \dots, \theta_t^n) \\
\dot{\theta}_t^{\,i} &= F_i(t, \theta_t^{i-1}, \dots, \theta_t^n), \quad 2 \le i \le n,
\end{aligned} \tag{1.4}$$

with the initial condition $\boldsymbol{\theta}_0(\mathbf{x}) = (\theta_0^1(\mathbf{x}), \dots, \theta_n^1(\mathbf{x})) = \mathbf{x}$. (Of course, (1.4) is the deterministic counterpart of (1.1).)

Comments. The transport of the initial condition $\mathbf{x}$ by the deterministic flow $\boldsymbol{\theta}$ in the density bounds may be seen as a nonlinear generalization of the Kolmogorov example. (See (1.3).) The i^{th} coordinate X_t^i of the system at time t oscillates around $\theta_t^i(\mathbf{x})$ with fluctuations of order $t^{i-1/2}$. For $i = 1$, the action of the flow in small time is negligible in comparison with the action of the noise: the order of the distance between $\theta^1(\mathbf{x})$ and x_1 is at most t, whereas the fluctuations of X_t^1 have $t^{1/2}$ as order. In short, this is the reason why the drift has no role in small time in the Aronson estimates. On the contrary, for $i \ge 2$, the fluctuations of X_t^i may be much less than the distance between $\theta^i(\mathbf{x})$ and x_i: the transport term in the density bounds has a key role in high coordinates.

Such a multiscale effect follows from the weak Hörmander setting we here consider and, more precisely, from the one-way propagation assumption of the noise. In comparison, it seems much more difficult to specify the right scales of each coordinate in the strong Hörmander framework investigated by Kusuoka and Stroock [KS87]: roughly speaking, Theorem 4.9 in [KS87] suggests that, at time t less than 1, the whole diffusion process (i.e. without any distinction between the coordinates) lives at scale $t^{1/2}$ (i.e. admits a diffusive scaling) with respect to the so-called *Carnot-Carathéodory distance* generated by the operator.

In Kusuoka and Stroock [KS87], the *Carnot-Carathéodory distance* yields the "off-diagonal" decay of the density. It is also referred as a *control metric* since it may be seen as the optimal cost of some deterministic control problem. As explained in Section 2 below, it turns out that this specific deterministic control problem is nothing but the deterministic counterpart of the stochastic control problem deriving from the Fleming transform (have in mind that the Fleming transform provides an explicit representation for $-\ln(p(T, \mathbf{x}, \mathbf{y}))$ in terms of some controlled diffusion process): this makes a formal connection between the *control metric* used in Kusuoka and Stroock [KS87] and the approach developed by Sheu [She91]. In our own setting, this formal connection is shown to be rigorous: the "off-diagonal" decay in Theorem 1.1 derives from some determistic control problem as well. Nevertheless, because of the flow $\boldsymbol{\theta}$, the decay has some complicated time-dynamics here and cannot be written in terms of some time-independent distance explicitly.

To finish with, we also indicate that the "diagonal" decay of the density has a structure similar to the one exhibited in Kusuoka and Stroock [KS87]: it may be explained as the sum of the lengths of the commutators spanning the whole space. (Here, the length of a commutator of some vector fields is equal to the sum of the inverses of the degrees of the underlying vector fields.) For example, for $d = 1$, the commutators used to span the whole space at some point $(t, \mathbf{x}) \in (0, +\infty) \times \mathbb{R}^n$, i.e. A, $[A, B]$, $[[A, B], B]$, $\dots$, with the same notations as in (1.2), have $1/2$, $3/2$, $\dots$, $(2n-1)/2$ as lengths [5], so that the sum is equal to $n(n-1)/2 + n/2 = n^2/2$. This matches the exponent of the "diagonal" decay of the density. Obviously, the argument also holds for $d \geq 2$.

Useful Notations. In what follows, we always denote by a bold letter a quantity in $\mathbb{R}^{nd}$: for example, zero in bold face, i.e. $\mathbf{0}$, stands for zero in $\mathbb{R}^{nd}$ and the solution $(X_t^1, \dots, X_t^n)_{t \geq 0}$ to (1.1) is denoted by $(\mathbf{X}_t)_{t \geq 0}$. Similary, Eq. (1.1) itself is written in a shortened form:

$$d\mathbf{X}_t = \mathbf{F}(t, \mathbf{X}_t) + B\sigma(t, \mathbf{X}_t)dW_t,$$

where $\mathbf{F} = (F_1, \dots, F_n)$ is an $\mathbb{R}^{nd}$-valued function and B stands for the $nd \times d$ matrix, which embeds $\mathbb{R}^d$ into $\mathbb{R}^{nd}$, i.e. $B = (I_d, 0, \dots, 0)^*$, where "$*$" stands for the transpose. Moreover, for $\mathbf{x} = (x_1, \dots, x_n) \in (\mathbb{R}^d)^n$, we set $\mathbf{x}^{i,n} = (x_i, \dots, x_n) \in (\mathbb{R}^d)^{n-i+1}$.

The frequently used expression "known parameters in $(\mathbf{A})$" refers to Λ, η, κ, $\mathcal{E}_1$, $\mathcal{E}_2$, $\dots$, $\mathcal{E}_n$, n and d. In particular, we emphasize that T is not considered as a parameter in $(\mathbf{A})$. This permits to make the distinction between the constants depending on T and the parameters in $(\mathbf{A})$ only and the constants depending on the parameters in $(\mathbf{A})$ only (and not on T). Except when specified, the constants mentioned below do not depend on other quantities.

Organization of the Paper. In Section 2 we specify the key steps to prove Theorem 1.1 and introduce the mathematical objects needed. It appears in

[5] Note that the vector field that generates the time component, i.e. the vector field B, is not taken into account to compute the diagonal decay at a given time.

that section that Gaussian linear systems and some related deterministic control problems play a central role in the whole proof. We give in Section 3 useful estimates on these objects. This is the technical core of the paper. Section 4 is dedicated to the proof of the lower bound. The upper bound is derived in Section 5. Some technical computations are postponed to Appendix A and B. Eventually, we give in Appendix C an alternative proof of the lower bound deriving from the parametrix expansion of the density and a suitable non standard chaining argument.

2 Strategy and Associated Mathematical Tools

We first discuss the existence of a transition density $(p(s,t,\mathbf{x},\mathbf{y}))_{0\le s<t;\mathbf{x},\mathbf{y}\in\mathbb{R}^{nd}}$ to the solution $(\mathbf{X}_t)_{t\ge 0}$ of (1.1). When the coefficients $\mathbf{F}$ and σ satisfy $(\mathbf{A})$ and are $\mathcal{C}^\infty$ in time and space, Hörmander's theorem for parabolic hypoellipticity (see [Hör67, Theorem 1.1]) applies: the transition density exists and is $\mathcal{C}^\infty$ in all the parameters. Moreover, it satisfies the Fokker–Planck equation:

$$\begin{aligned}&\partial_t p(t,T,\mathbf{x},\mathbf{y}) + \mathcal{L}_{t,\mathbf{x}} p(t,T,\mathbf{x},\mathbf{y}) = 0, \quad 0\le t<T,\ \mathbf{x},\mathbf{y}\in\mathbb{R}^{nd},\\ &p(T,T,\mathbf{x},\mathbf{y}) = \delta_{\mathbf{y}}(\mathbf{x}),\ \mathbf{x},\mathbf{y}\in\mathbb{R}^{nd},\end{aligned} \tag{2.1}$$

where $\mathcal{L}_{t,\mathbf{x}} = (1/2)\mathrm{Tr}(a(t,\mathbf{x})D^2_{x_1}) + F_1(t,\mathbf{x})D_{x_1} + \sum_{i=2}^n F_i(t,\mathbf{x}^{i-1,n})D_{x_i}$ is the infinitesimal generator of $\mathbf{X}$ at time t.

To recover this framework, we assume below that the coefficients are smooth. By a regularization procedure, this is always possible: we can find a mollifying sequence of $(\mathbf{F},\sigma)$ satisfying $(\mathbf{A})$ uniformly. We emphasize that the density bounds we exhibit below in the regularized setting depends on T and the quantities in $(\mathbf{A})$ only and are independent of the regularization procedure. Theorem 1.1, under $(\mathbf{A})$ only, then follows by letting the regularization parameter tend to zero and by using two key arguments: uniqueness in law for (1.1) and Radon–Nikodym's theorem.

The Fokker–Planck equation is the starting point of our approach. By chaining (2.1) with the function $-\ln$, we indeed deduce a probabilistic representation of the transition density: for $0\le t<T$, $-\ln(p(t,T,\mathbf{x},\mathbf{y}))$ is shown to be the value function of some optimal stochastic control problem. This procedure is known as Fleming's logarithmic transform. It has been used by Sheu [She91] to recover the Aronson bounds for uniformly elliptic diffusion processes. We here adapt it to our framework.

Before we explain what Fleming's transform is, we emphasize that $p(0,T,\cdot,\cdot)$ is often denoted by $p(T,\cdot,\cdot)$ in the paper. This is the notation used in the statement of Theorem 1.1.

We now fix the arrival point $(T, \mathbf{y}_0) \in (0, +\infty) \times \mathbb{R}^{nd}$ as in the statement of Theorem 1.1. To explain Fleming's transform, we first approximate the Dirac boundary condition in (2.1) by a true function: we here introduce a sequence $(\eta_\varepsilon)_{\varepsilon>0}$ of mollifiers on $\mathbb{R}^{nd}$ weakly converging towards the Dirac mass $\delta_{\mathbf{y}_0}$ at $\mathbf{y}_0$. In addition, we assume that the $(\eta_\varepsilon)_{\varepsilon>0}$ are positive (i.e. > 0) on the whole $\mathbb{R}^{nd}$ and satisfy

$$\exists \varepsilon_0 > 0 : \lim_{c \to +\infty} \sup_{0<\varepsilon<\varepsilon_0} \sup_{|\mathbf{y}|>c} \eta_\varepsilon(\mathbf{y}) = 0. \tag{2.2}$$

(We provide below an example for such mollifiers.) We then approximate the transition density by setting, for all $\varepsilon > 0$ and $(t, \mathbf{x}) \in [0, T-\varepsilon] \times \mathbb{R}^{nd}$, $u_\varepsilon(t, \mathbf{x}) = \mathbb{E}[\eta_\varepsilon(\mathbf{X}^{t,\mathbf{x}}_{T-\varepsilon})]$. (Here, the superscript $(t, \mathbf{x})$ in the notation $\mathbf{X}^{t,\mathbf{x}}$ means that $\mathbf{X}_t = \mathbf{x}$ in (1.1).) The coefficients of $\mathbf{X}$ being smooth, it solves in the classical sense the Cauchy problem $\partial_t u_\varepsilon(t, \mathbf{x}) + \mathcal{L}_t u_\varepsilon(t, \mathbf{x}) = 0$, $0 \le t < T-\varepsilon$, $\mathbf{x} \in \mathbb{R}^{nd}$, with the boundary condition $u_\varepsilon(T-\varepsilon, \mathbf{x}) = \eta_\varepsilon(\mathbf{x})$, $\mathbf{x} \in \mathbb{R}^{nd}$. (When possible, we remove the index $\mathbf{x}$ in $\mathcal{L}_{t,\mathbf{x}}$ introduced after (2.1).) Since p is continuous away from the boundary, we deduce from the localization property (2.2): $\lim_{\varepsilon\to 0} u_\varepsilon(0, \mathbf{x}) = \lim_{\varepsilon\to 0} \mathbb{E}[\eta_\varepsilon(\mathbf{X}^{0,\mathbf{x}}_{T-\varepsilon})] = \lim_{\varepsilon\to 0} \int_{\mathbb{R}^{nd}} \eta_\varepsilon(\mathbf{y}) p(0, T-\varepsilon, \mathbf{x}, \mathbf{y}) d\mathbf{y} = p(0, T, \mathbf{x}, \mathbf{y}_0) = p(T, \mathbf{x}, \mathbf{y}_0)$.

As announced above, we now set, for all $0 \le t \le T-\varepsilon$ and $\mathbf{x} \in \mathbb{R}^{nd}$, $J_\varepsilon(t, \mathbf{x}) = -\ln[u_\varepsilon(t, \mathbf{x})]$. (We emphasize that $u_\varepsilon(t, \mathbf{x}) > 0$ since η_ε is positive on the whole space.) We have $D_{x_i}(\text{resp. } \partial_t) J_\varepsilon(t, \mathbf{x}) = -u_\varepsilon^{-1}(t, \mathbf{x}) D_{x_i}(\text{resp. } \partial_t) u_\varepsilon(t, \mathbf{x})$ and $D^2_{x_1,x_1} J_\varepsilon(t, \mathbf{x}) = -u_\varepsilon^{-1}(t, \mathbf{x}) D^2_{x_1,x_1} u_\varepsilon(t, \mathbf{x}) + u_\varepsilon^{-2}(t, \mathbf{x}) D_{x_1} u_\varepsilon(t, \mathbf{x}) \otimes D_{x_1} u_\varepsilon(t, \mathbf{x}) = -u_\varepsilon^{-1}(t, \mathbf{x}) D^2_{x_1,x_1} u_\varepsilon(t, \mathbf{x}) + D_{x_1} J_\varepsilon(t, \mathbf{x}) \otimes D_{x_1} J_\varepsilon(t, \mathbf{x})$. We deduce that J_ε satisfies the following nonlinear parabolic equation with quadratic growth

$$\partial_t J_\varepsilon(t, \mathbf{x}) + \mathcal{L}_t J_\varepsilon(t, \mathbf{x}) - \frac{1}{2} \langle a(t, \mathbf{x}) D_{x_1} J_\varepsilon(t, \mathbf{x}), D_{x_1} J_\varepsilon(t, \mathbf{x}) \rangle = 0, \tag{2.3}$$

$0 \le t < T - \varepsilon$, $\mathbf{x} \in \mathbb{R}^{nd}$, with the boundary condition $J_\varepsilon(T-\varepsilon, \mathbf{x}) = -\ln(\eta_\varepsilon(\mathbf{x}))$, $\mathbf{x} \in \mathbb{R}^{nd}$. The key point is now to notice that the quadratic part of the above equation can be rewritten as

$$\partial_t J_\varepsilon(t, \mathbf{x}) + \mathcal{L}_t J_\varepsilon(t, \mathbf{x}) + \inf_{v \in \mathbb{R}^d} \Big[\langle v, D_{x_1} J_\varepsilon(t, \mathbf{x}) \rangle + \frac{1}{2} \langle a^{-1}(t, \mathbf{x}) v, v \rangle \Big] = 0. \tag{2.4}$$

For a given $(t, \mathbf{x}) \in [0, T-\varepsilon] \times \mathbb{R}^{nd}$, the above infimum is reached at $v = v(t, \mathbf{x}) := -a(t, \mathbf{x}) D_{x_1} J_\varepsilon(t, \mathbf{x})$ which indeed gives (2.3). Denoting by $\mathcal{P}(T-\varepsilon)$ the set of progressively measurable processes $(v_t)_{0 \le t \le T-\varepsilon}$ with values in $\mathbb{R}^d$ s.t. $\mathbb{E}[\int_0^{T-\varepsilon} |v_t|^2 dt] < +\infty$, we can associate with each $(v_t)_{0 \le t \le T-\varepsilon} \in \mathcal{P}(T-\varepsilon)$ the controlled version of (1.1), i.e.

$$d\boldsymbol{\chi}_t = [\mathbf{F}(t, \boldsymbol{\chi}_t) + B v_t]\, dt + B\sigma(t, \boldsymbol{\chi}_t) dW_t,\ \boldsymbol{\chi}_0 = \mathbf{x}. \tag{2.5}$$

Following (2.4), we can write $J_\varepsilon(0,\mathbf{x})$ as the value function of some stochastic optimization problem:

$$J_\varepsilon(0,\mathbf{x}) = \inf_{(v_t)_t \in \mathcal{P}(T-\varepsilon)} \mathbb{E}\left[\frac{1}{2}\int_0^{T-\varepsilon} \langle a^{-1}(t,\boldsymbol{\chi}_t^{0,\mathbf{x}})v_t, v_t\rangle dt - \ln[\eta_\varepsilon(\boldsymbol{\chi}_{T-\varepsilon}^{0,\mathbf{x}})]\right]. \quad (2.6)$$

(The superscript $(0,\mathbf{x})$ in $\boldsymbol{\chi}^{0,\mathbf{x}}$ means that $\boldsymbol{\chi}_0 = \mathbf{x}$.) Precisely, Eq. (2.6) follows from Itô's formula. Indeed, expanding $(J_\varepsilon(t,\boldsymbol{\chi}_t))_{0\le t\le T-\varepsilon}$ (to simplify, we here remove the superscript $(0,\mathbf{x})$), we obtain from (2.3)

$$\begin{aligned}
dJ_\varepsilon(t,\boldsymbol{\chi}_t) &= \big[\partial_t J_\varepsilon(t,\boldsymbol{\chi}_t) + \mathcal{L}_t J_\varepsilon(t,\boldsymbol{\chi}_t) + \langle D_{x_1}J_\varepsilon(t,\boldsymbol{\chi}_t), v_t\rangle\big]dt \\
&\quad + \langle D_{x_1}J_\varepsilon(t,\boldsymbol{\chi}_t), \sigma(t,\boldsymbol{\chi}_t)dW_t\rangle \\
&= \Big[\frac{1}{2}\langle a^{-1}(t,\boldsymbol{\chi}_t)v_t^*, v_t^*\rangle - \langle a^{-1}(t,\boldsymbol{\chi}_t)v_t, v_t^*\rangle\Big]dt - \langle \sigma^{-1}(t,\boldsymbol{\chi}_t)v_t^*, dW_t\rangle \\
&= \frac{1}{2}\Big[|\sigma^{-1}(t,\boldsymbol{\chi}_t)[v_t^* - v_t]|^2 - |\sigma^{-1}(t,\boldsymbol{\chi}_t)v_t|^2\Big]dt - \langle \sigma^{-1}(t,\boldsymbol{\chi}_t)v_t^*, dW_t\rangle,
\end{aligned}$$

where $v_t^* = -a(t,\boldsymbol{\chi}_t)D_{x_1}J_\varepsilon(t,\boldsymbol{\chi}_t)$, so that

$$\begin{aligned}
J_\varepsilon(0,\mathbf{x}) &= -\ln[\eta_\varepsilon(\boldsymbol{\chi}_{T-\varepsilon})] + \frac{1}{2}\int_0^{T-\varepsilon} \langle a^{-1}(t,\boldsymbol{\chi}_t)v_t, v_t\rangle dt \\
&\quad - \frac{1}{2}\int_0^{T-\varepsilon} |\sigma^{-1}(t,\boldsymbol{\chi}_t)[v_t^* - v_t]|^2 dt + \int_0^{T-\varepsilon} \langle \sigma^{-1}(t,\boldsymbol{\chi}_t)v_t^*, dW_t\rangle.
\end{aligned} \quad (2.7)$$

Eq. (2.6) together with the limit $-\ln(p(T,\mathbf{x},\mathbf{y}_0)) = \lim_{\varepsilon\to 0} J_\varepsilon(0,\mathbf{x})$ is the "Master formula" in our proof. The proof for both bounds will rely on it.

2.2 Steps for the Lower Bound

Here is the key idea: to obtain a lower bound for $p(T,\mathbf{x}_0,\mathbf{y}_0)$, $\mathbf{x}_0 \in \mathbb{R}^{nd}$, it is sufficient to obtain an upper bound for $J_\varepsilon(0,\mathbf{x}_0)$, uniformly in $\varepsilon > 0$. Since $J_\varepsilon(0,\mathbf{x}_0)$ is the value function of the minimization problem (2.6), the whole problem thus consists in finding a relevant control process $(v_t)_{0\le t\le T-\varepsilon}$ for each $\varepsilon > 0$.

The choice for $(v_t)_{0\le t\le T-\varepsilon}$ is performed in three steps:

(i). First, we investigate a reference case, namely the linear one, to understand how things work. Indeed, in the linear case, i.e. when the drift $\mathbf{F}$ is affine in space (and satisfies $(\mathbf{A})$) and the matrix σ is constant in space, the solution $(\mathbf{X}_t)_{t\ge 0}$ to (1.1) is a Gaussian process: under $(\mathbf{A})$, the transition density exists and has an explicit form. Moreover, the optimal control in (2.5–2.6) can also be written in an explicit way. This preliminary analysis is performed in Section 3.

(ii). To recover at best the linear (or Gaussian) case, the strategy consists in linearizing the controlled equation (2.5). The point is then to choose a (deterministic) curve $(\boldsymbol{\phi}_t)_{0\le t\le T}$ to linearize around, i.e. to expand $\mathbf{F}(t,\boldsymbol{\chi}_t^{0,\mathbf{x}_0})$ in (2.5) as $\mathbf{F}(t,\boldsymbol{\chi}_t^{0,\mathbf{x}_0}) = \mathbf{F}(t,\boldsymbol{\phi}_t)+\mathbf{D_x F}(t,\boldsymbol{\phi}_t)(\boldsymbol{\chi}_t^{0,\mathbf{x}_0}-\boldsymbol{\phi}_t)+o(|\boldsymbol{\chi}_t^{0,\mathbf{x}_0}-\boldsymbol{\phi}_t|)$, where $\mathbf{D_x F}\in\mathcal{M}_{nd}(\mathbb{R})$ is the space derivative of $\mathbf{F}$. (And similarly, to approximate $\sigma(t,\boldsymbol{\chi}_t^{0,\mathbf{x}_0})$ by $\sigma(t,\boldsymbol{\phi}_t)$.) Since $(\boldsymbol{\phi}_t)_{0\le t\le T}$ is deterministic, the mapping $\mathbf{z}\in\mathbb{R}^{nd}\mapsto \mathbf{F}(t,\boldsymbol{\phi}_t)+\mathbf{D_x F}(t,\boldsymbol{\phi}_t)(\mathbf{z}-\boldsymbol{\phi}_t)$ is indeed affine and the diffusion coefficient $(\sigma(t,\boldsymbol{\phi}_t))_{0\le t\le T}$ is deterministic, as required in (i). A natural choice is then to pick $(\boldsymbol{\phi}_t)_{0\le t\le T}$ as the solution of a deterministic version of the stochastic control problem (2.5), i.e. as the solution of the controlled ODE:

$$\dot{\boldsymbol{\phi}}_t = \mathbf{F}(t,\boldsymbol{\phi}_t) + B\varphi_t, \quad 0\le t\le T; \quad \boldsymbol{\phi}_0 = \mathbf{x}_0, \tag{2.8}$$

$(\varphi_t)_{0\le t\le T}$ standing for a deterministic control in $L^2([0,T],\mathbb{R}^d)$. The control φ is then chosen to make $\boldsymbol{\phi}$ hit $\mathbf{y}_0$ at time T, i.e. $\boldsymbol{\phi}_T=\mathbf{y}_0$, with the lowest possible quadratic cost, exactly as $(v_t)_{0\le t\le T-\varepsilon}$ in (2.6) has to be chosen to make $\boldsymbol{\chi}_{T-\varepsilon}^{0,\mathbf{x}_0}$ tend to $\mathbf{y}_0$ with ε with the lowest possible quadratic cost. In other words, we choose $(\varphi_t)_{0\le t\le T}$ as an optimal control for the minimization problem:

$$I(T,\mathbf{x}_0,\mathbf{y}_0) := \inf\Big\{\int_0^T |\varphi_t|^2 dt : \boldsymbol{\phi}_0=\mathbf{x}_0,\ \boldsymbol{\phi}_T=\mathbf{y}_0\Big\}. \tag{2.9}$$

In Section 3, we prove that $I(T,\mathbf{x}_0,\mathbf{y}_0)$ is indeed finite (Eq. (2.8) is then said to be controllable). We also prove that $I(T,\mathbf{x}_0,\mathbf{y}_0)$ is of the same order as $T|\mathbb{T}_T^{-1}(\boldsymbol{\theta}_T(\mathbf{x}_0)-\mathbf{y}_0)|^2$: comparing with the statement of Theorem 1.1, we thus understand that the cost of φ corresponds to the "off-diagonal" decay of the density. The linearization procedure is detailed in Section 4.

(iii). Once $(\varphi_t)_{0\le t\le T}$ has been chosen, it remains to choose $(v_t-\varphi_t)_{0\le t\le T}$. By (2.5) and (2.8), we observe that $(\boldsymbol{\chi}_t^{0,\mathbf{x}}-\boldsymbol{\phi}_t)_{0\le t\le T-\varepsilon}$ satisfies the linearized equation

$$d\big[\boldsymbol{\chi}_t^{0,\mathbf{x}}-\boldsymbol{\phi}_t\big] = \big[\mathbf{D_x F}(t,\boldsymbol{\phi}_t)(\boldsymbol{\chi}_t^{0,\mathbf{x}}-\boldsymbol{\phi}_t)+B(v_t-\varphi_t)\big]dt+\sigma(t,\boldsymbol{\phi}_t)dW_t+dR_t,$$

$(R_t)_{0\le t\le T-\varepsilon}$ denoting a remaining term that is expected to be well-controlled. We then notice that the starting point $\boldsymbol{\chi}_0^{0,\mathbf{x}}-\boldsymbol{\phi}_0$ is $\mathbf{0}$ and that the final point $\boldsymbol{\chi}_{T-\varepsilon}^{0,\mathbf{x}}-\phi_{T-\varepsilon}$ is also expected to be close to $\mathbf{y}_0-\mathbf{y}_0=\mathbf{0}$. In some sense, we are reduced to the linear case (i) with $\mathbf{0}$ and $\mathbf{0}$ as boundary points and with $(v_t-\varphi_t)_{0\le t\le T-\varepsilon}$ as control: we then know how to perform a relevant choice for $(v_t-\varphi_t)_{0\le t\le T-\varepsilon}$, or equivalently for $(v_t)_{0\le t\le T-\varepsilon}$. By Step (i), we also know the associated cost (in the sense of (2.6)): we will see in Section 3 that it is equal to $-\ln(T^{-n^2d/2})$ up to an additive constant. This corresponds in Theorem 1.1 to the "diagonal" decay of the density.

Remark 2.1 *The function $I(T,\mathbf{x},\mathbf{y})$ is known as the action functional in large deviations theory. It provides in short time a natural link between the deterministic control problem (2.8–2.9) and the transition density of* (1.1), *see e.g. Freidlin and Wentzell [FW98] for the non degenerate case and Ben Arous and Léandre [BL91] for results under the strong Hörmander condition. The technique introduced by Sheu [She91] and here developed may be seen as an extension of this connection when time is not necessarily small.*

2.3 Intrinsic Scaling Property of the System

The reader may object that the previous arguments only permit to get a lower bound for the density at time T, whereas the lower bound in Theorem 1.1 is given for any $t \in (0, T]$, the underlying constant C_T depending on T (and on $(\mathbf{A})$) only and not on t. A possible way to pass from t to T consists in using intrinsic scaling properties of the system (1.1).

To simplify, we here explain how to pass from T to 1. For a given $T > 0$, we define the rescaled version of $(\mathbf{X}_t)_{t\geq 0}$ by setting

$$\begin{aligned}\hat{\mathbf{X}}_t = (\hat{X}_t^1, \dots, \hat{X}_t^n)^* &= T^{1/2}\mathbb{T}_T^{-1}\mathbf{X}_{Tt} \\ &= (T^{-1/2}X_{Tt}^1, T^{-3/2}X_{Tt}^2, \dots, T^{-(2n-1)/2}X_{Tt}^n)^*, \quad t \geq 0,\end{aligned}$$

where $\mathbb{T}_T$ stands for the scale matrix of the system at time T, as defined in the statement of Theorem 1.1. The rescaled process $(\hat{\mathbf{X}}_t)_{t\geq 0}$ then satisfies:

$$d\hat{\mathbf{X}}_t = T^{3/2}\mathbb{T}_T^{-1}\mathbf{F}(Tt, T^{-1/2}\mathbb{T}_T\hat{\mathbf{X}}_t)dt + B\sigma(Tt, T^{-1/2}\mathbb{T}_T\hat{\mathbf{X}}_t)d\hat{W}_t, \qquad (2.10)$$

$\hat{W}$ standing for the rescaled Brownian motion $(\hat{W}_t = T^{-1/2}W_{Tt})_{t\geq 0}$. Setting

$$\hat{\sigma}(t, \mathbf{x}) = \sigma(Tt, T^{-1/2}\mathbb{T}_T\mathbf{x}),\ \hat{\mathbf{F}}(t, \mathbf{x}) = T^{3/2}\mathbb{T}_T^{-1}\mathbf{F}(Tt, T^{-1/2}\mathbb{T}_T\mathbf{x}), \quad t \geq 0,$$

we can see the system (2.10) as a system of type (1.1) with $\sigma, \mathbf{F}$ replaced by $\hat{\sigma}, \hat{\mathbf{F}}$ and W replaced by the rescaled Brownian motion $\hat{W}$. For $T \leq 1$, the coefficients $\hat{\sigma}$ and $\hat{\mathbf{F}}$ satisfy the same assumptions as σ and $\mathbf{F}$, i.e. with the same constants. For $T > 1$, $\hat{\sigma}$ and $\hat{\mathbf{F}}$ still satisfy the same type of assumptions as σ and $\mathbf{F}$ but with constants magnified by a power of T, namely T^n. In short, $D_{\mathbf{x}_i}\hat{\mathbf{F}}_j(t, \mathbf{x}) = T^{1-j+i}D_{\mathbf{x}_i}\mathbf{F}_j(Tt, T^{-1/2}\mathbb{T}_T\mathbf{x})$ for $\mathbf{x} \in \mathbb{R}^{nd}$. Since $D_{\mathbf{x}_i}\mathbf{F}_j = 0$ for $i < j - 1$, this completes the proof. Note also that under $(\mathbf{A})$, $D_{\mathbf{x}_{i-1}}\hat{\mathbf{F}}_i(t, \mathbf{x}) = D_{\mathbf{x}_{i-1}}\mathbf{F}_i(Tt, T^{-1/2}\mathbb{T}_T\mathbf{x}) \in \mathcal{E}_{i-1}, 2 \leq i \leq n$.

Denoting by $(\hat{p}(1, \mathbf{x}, \mathbf{y}))_{\mathbf{x},\mathbf{y}\in\mathbb{R}^{nd}}$ the transition density of $\hat{\mathbf{X}}$ at time 1, we have $\hat{p}(1, \mathbf{x}, \mathbf{y}) = T^{n^2d/2}p(T, T^{-1/2}\mathbb{T}_T\mathbf{x}, T^{-1/2}\mathbb{T}_T\mathbf{y})$ by change of variable (since $\det(T^{-1/2}\mathbb{T}_T) = T^{n^2d/2}$), i.e.

$$\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^{nd}, \quad p(T, \mathbf{x}, \mathbf{y}) = T^{-n^2d/2}\hat{p}(1, T^{1/2}\mathbb{T}_T^{-1}\mathbf{x}, T^{1/2}\mathbb{T}_T^{-1}\mathbf{y}). \qquad (2.11)$$

Eq. (2.11) shows how to deduce an estimate for the transition density at time T from an estimate at time 1. Assume indeed that one of the two bounds in

Theorem 1.1 holds at time 1. Then, $\hat{p}(1,\mathbf{x},\mathbf{y})$, $\mathbf{x},\mathbf{y}\in\mathbb{R}^{nd}$, admits the same bound, i.e. $C\exp(-C^{-1}|\hat{\boldsymbol{\theta}}_1(\mathbf{x})-\mathbf{y}|^2)$ (with $C\geq 1$ for the upper bound, $C\leq 1$ for the lower bound), the constant C possibly depending on T when $T\geq 1$. Here, $(\hat{\boldsymbol{\theta}}_t(\mathbf{x}))_{0\leq t\leq 1}$ stands for the rescaled flow, initialized at $\mathbf{x}$, i.e. the solution of the ODE (1.4), but driven by $\hat{\mathbf{F}}$. Following (2.10), it is plain to see that

$$\hat{\boldsymbol{\theta}}_t(\mathbf{x}) = T^{1/2}\mathbb{T}_T^{-1}\boldsymbol{\theta}_{Tt}(T^{-1/2}\mathbb{T}_T\mathbf{x}), \quad (t,\mathbf{x})\in[0,1]\times\mathbb{R}^{nd}. \tag{2.12}$$

Plugging (2.12) into the bound $C\exp(-C^{-1}|\hat{\boldsymbol{\theta}}_1(\mathbf{x})-\mathbf{y}|^2)$, we deduce from (2.11) that $p(T,\mathbf{x},\mathbf{y})$ admits $CT^{-n^2d/2}\exp(-C^{-1}T|\mathbb{T}_T^{-1}(\boldsymbol{\theta}_T(\mathbf{x})-\mathbf{y})|^2)$ as bound.

We here provide another application of (2.12). We indeed emphasize that $\hat{\boldsymbol{\theta}}_1$ is a diffeomorphism on $\mathbb{R}^{nd}$ with a Lipschitz converse. (For $\mathbf{y}\in\mathbb{R}^{nd}$, $\hat{\boldsymbol{\theta}}_1^{-1}(\mathbf{y})$ is the value at time 0 of the solution matching $\mathbf{y}$ at time 1 to the ODE (1.4) driven by $\hat{\mathbf{F}}$.) Denoting by C the "bi-Lipschitz" constant of $\hat{\boldsymbol{\theta}}_1$, we obtain

$$\forall \mathbf{x},\mathbf{y}\in\mathbb{R}^{nd}, \quad C^{-1}|\mathbf{x}-\hat{\boldsymbol{\theta}}_1^{-1}(\mathbf{y})|^2 \leq |\hat{\boldsymbol{\theta}}_1(\mathbf{x})-\mathbf{y}|^2 \leq C|\mathbf{x}-\hat{\boldsymbol{\theta}}_1^{-1}(\mathbf{y})|^2.$$

Changing $\mathbf{x}$ into $T^{1/2}\mathbb{T}_T^{-1}\mathbf{x}$ and $\mathbf{y}$ into $T^{1/2}\mathbb{T}_T^{-1}\mathbf{y}$ and plugging (2.12), we deduce that for all $\mathbf{x},\mathbf{y}\in\mathbb{R}^{nd}$

$$C^{-1}\Big|\mathbb{T}_T^{-1}\big[\mathbf{x}-\boldsymbol{\theta}_T^{-1}(\mathbf{y})\big]\Big|^2 \leq \Big|\mathbb{T}_T^{-1}\big[\boldsymbol{\theta}_T(\mathbf{x})-\mathbf{y}\big]\Big|^2 \leq C\Big|\mathbb{T}_T^{-1}\big[\mathbf{x}-\boldsymbol{\theta}_T^{-1}(\mathbf{y})\big]\Big|^2. \tag{2.13}$$

The constant C in (2.13) is related to the rescaled coefficient $\hat{\mathbf{F}}$ only: in particular, C is independent of T for $T\leq 1$. Plugging (2.13) into the statement of Theorem 1.1, we understand that the off-diagonal bound for the density may be obtained by transporting the initial condition $\mathbf{x}$ by the forward flow $\boldsymbol{\theta}_T$ or by transporting the terminal condition $\mathbf{y}$ by the backward flow $\boldsymbol{\theta}_T^{-1}$.

2.4 Steps for the Upper Bound

The proof of the upper bound for the density both follows from the Fleming representation (2.6) and from the McKean–Singer parametrix method, see [MS67]. The coupling of these two arguments seems completely new in the literature. In fact, in some specific cases, see e.g. [KM00] and [KMM09], the parametrix method has already been applied to obtain an upper bound for the transition density of (1.1), but separately from any stochastic control argument. The framework we here deal with seems too general to repeat the proofs given in these two papers: the Fleming representation (2.6) then provides supplementary material that permits to make the whole thing work.

The McKean–Singer expansion is an expansion of the density of a stochastic system in terms of the Gaussian kernel of some related linear stochastic system. As the strategy used for the lower bound, it is a perturbation method: the coefficients of the Gaussian system are obtained by linearization. We here explain the general principle of the parametrix representation of $p(T, \mathbf{x}_0, \mathbf{y}_0)$.

(i). The first step consists in approximating the density $p(T, \cdot, \mathbf{y}_0)$ by a suitable known Gaussian density $\tilde{p}^{T,\mathbf{y}_0}$: the superscript $(T, \mathbf{y}_0)$ in $\tilde{p}^{T,\mathbf{y}_0}$ indicates the possible dependence of the mean and the covariance of $\tilde{p}$ on the boundary conditions $(T, \mathbf{y}_0)$. The choice obeys the following intuitive idea: in short time, $p(T, \cdot, \mathbf{y}_0)$ and $\tilde{p}^{T,\mathbf{y}_0}$ are to be close.

(ii). As in point (ii) of the previous subsection, the Gaussian system to consider is obtained by linearization of (1.1). (Here, the equation to linearize is (1.1) and not (2.5).) Again, the question is to choose a deterministic curve to linearize around. Since the key role is here played by the arrival point $\mathbf{y}_0$, the right path to consider is the solution matching $\mathbf{y}_0$ at time T of the ODE driven by $\mathbf{F}$, that is $(\boldsymbol{\theta}_{t,T}(\mathbf{y}_0))_{0 \le t \le T}$, solution of

$$\frac{d}{dt}\boldsymbol{\theta}_{t,T}(\mathbf{y}_0) = \mathbf{F}\Big(t, \boldsymbol{\theta}_{t,T}(\mathbf{y}_0)\Big), \quad 0 \le t \le T; \quad \boldsymbol{\theta}_{T,T}(\mathbf{y}_0) = \mathbf{y}_0. \tag{2.14}$$

(iii). We thus introduce the Gaussian system

$$\begin{aligned} d\tilde{\mathbf{X}}_t = &\Big[\mathbf{F}(t, \boldsymbol{\theta}_{t,T}(\mathbf{y}_0)) + \mathbf{D}_{\mathbf{x}}\mathbf{F}(t, \boldsymbol{\theta}_{t,T}(\mathbf{y}_0))\Big(\tilde{\mathbf{X}}_t - \boldsymbol{\theta}_{t,T}(\mathbf{y}_0)\Big)\Big]dt \\ &+ B\sigma(t, \boldsymbol{\theta}_{t,T}(\mathbf{y}_0))dW_t, \quad 0 \le t \le T, \end{aligned} \tag{2.15}$$

obtained by linearization of the dynamics of $\mathbf{X}$. We prove in Section 3 that $(\tilde{\mathbf{X}}_t)_{t\in[0,T]}$ has a transition density $(\tilde{p}^{T,\mathbf{y}_0}(s, t, \mathbf{x}, \mathbf{y}))_{0 \le s < t \le T; \mathbf{x}, \mathbf{y} \in \mathbb{R}^{nd}}$ under $(\mathbf{A})$. Since the notation $\tilde{p}^{T,\mathbf{y}_0}(s, t, \mathbf{x}, \mathbf{y})$ is a bit heavy to handle, we adopt the following convention: $\tilde{p}(s, T, \mathbf{x}, \mathbf{y}_0)$ stands for $\tilde{p}^{T,\mathbf{y}_0}(s, T, \mathbf{x}, \mathbf{y}_0)$. (That is: we forget the superscript when the terminal point at which the density is evaluated coincides with the "freezing point" in (2.15).)

Denoting by $(\tilde{\mathcal{L}}_t^{T,\mathbf{y}_0})_{0 \le t \le T}$ the generator of the process $\tilde{\mathbf{X}}$ frozen at time T and at point $\mathbf{y}_0$, we recall from [MS67] that the key quantity in the parametrix method is the kernel

$$\begin{aligned} H(t, T, \mathbf{x}, \mathbf{y}_0) &= \Big[\mathcal{L}_{t,\mathbf{x}} - \tilde{\mathcal{L}}_{t,\mathbf{x}}^{T,\mathbf{y}_0}\Big]\Big(\tilde{p}^{T,\mathbf{y}_0}(t, T, \mathbf{x}, \mathbf{y}_0)\Big) \\ &\Big\{ = \Big[\mathcal{L}_{t,\mathbf{x}} - \tilde{\mathcal{L}}_{t,\mathbf{x}}^{T,\mathbf{y}_0}\Big]\Big(\tilde{p}(t, T, \mathbf{x}, \mathbf{y}_0)\Big)\Big\}, \end{aligned} \tag{2.16}$$

where $(\mathcal{L}_t)_{0 \le t \le T}$ stands for the original differential operator associated with $\mathbf{F}$ and σ. (See (2.1).) Using the Kolmogorov equations, we indeed write:

$$p(T, \mathbf{x}_0, \mathbf{y}_0) = \tilde{p}(0, T, \mathbf{x}_0, \mathbf{y}_0) + \int_0^T \int_{\mathbb{R}^{nd}} p(t, \mathbf{x}_0, \mathbf{z}) H(t, T, \mathbf{z}, \mathbf{y}_0) dt d\mathbf{z}.$$

The idea of the McKean–Singer parametrix consists in iterating this representation formula. An induction shows that, for any integer $N \geq 1$,

$$p(T, \mathbf{x}_0, \mathbf{y}_0) = \tilde{p}(0, T, \mathbf{x}_0, \mathbf{y}_0) + \sum_{k=1}^{N} \int_0^T \int_{\mathbb{R}^{nd}} \tilde{p}(0, t, \mathbf{x}_0, \mathbf{z}) H^{\otimes k}(t, T, \mathbf{z}, \mathbf{y}_0) dt d\mathbf{z}$$
$$+ \int_0^T \int_{\mathbb{R}^{nd}} p(t, \mathbf{x}_0, \mathbf{z}) H^{\otimes (N+1)}(t, T, \mathbf{z}, \mathbf{y}_0) dt d\mathbf{z}. \tag{2.17}$$

Again, $\tilde{p}(0, t, \mathbf{x}_0, \mathbf{z})$ stands for $\tilde{p}^{t,\mathbf{z}}(0, t, \mathbf{x}_0, \mathbf{z})$. (The freezing point is $(t, \mathbf{z})$.) Moreover,

$$H^{\otimes (k+1)}(t, T, \mathbf{z}, \mathbf{y}_0) := \int_t^T \int_{\mathbb{R}^{nd}} H(t, s, \mathbf{z}, \mathbf{z}') H^{\otimes k}(s, T, \mathbf{z}', \mathbf{y}_0) ds d\mathbf{z}'. \tag{2.18}$$

The standard argument in the McKean–Singer theory consists in letting N tend to the infinity to obtain a representation of p as an infinite sum of known convolution kernels. Unfortunately, this argument doesn't apply in our general framework: because of the transport $\boldsymbol{\theta}$, we fail to control the iterated kernels $(H^{\otimes k})_{k \geq 1}$ uniformly in k. (Again, we refer to [KMM09] for a specific type of Eq. (1.1) for which it works.) We thus need to truncate the series.

(iv). Here is the key point : by the stochastic control representation (2.6), we also manage to prove that, for N large enough, the remaining term

$$\int_0^T \int_{\mathbb{R}^{nd}} p(t, \mathbf{x}_0, \mathbf{z}) H^{\otimes (N+1)}(t, T, \mathbf{z}, \mathbf{y}_0) dt d\mathbf{z}$$

admits an upper Gaussian bound. This is sufficient to make the McKean–Singer approach work and also seems, to our best knowledge, to be new.

3 Gaussian Systems and Controllability

As emphasized in Section 2, Gaussian (or equivalently linear) systems play a key role in the proof of Theorem 1.1. They may be seen as the central example. In this section, we thus collect several results for Gaussian systems and relate them to deterministic controllability properties of (2.8)-(2.9).

3.1 Linear Systems

We here consider a measurable $\mathcal{M}_d(\mathbb{R})$-valued family $(\Sigma_t)_{0 \leq t \leq T}$ and a measurable $\mathcal{M}_{nd}(\mathbb{R})$-valued family $(\mathbf{L}_t)_{0 \leq t \leq T}$, T being positive as above, satisfying assumption

($\mathbf{A}^{\text{linear}}$) The spectrum of $A_t = \Sigma_t \Sigma_t^*$ is included in $[\Lambda^{-1}, \Lambda]$ for all $t \in [0, T]$. (The same Λ as in ($\mathbf{A}$).) Each $\mathbf{L}_t$, $0 \leq t \leq T$, is bounded by κ (the same as in ($\mathbf{A}$)) and is of the same form as $\mathbf{F}$ in (1.1): $\mathbf{L}_t$ may decomposed in $n \times n$ blocks of size $d \times d$, denoted by $([L_t]_{i,j})_{1 \leq i,j \leq n}$; for $1 \leq i, j \leq n$, the block $[L_t]_{i,j} = 0$ if $i \geq j + 2$ (the blocks under the subdiagonal are null: see the picture in footnote[6]). As in the picture in footnote, the blocks on the subdiagonal have a key role in the following: for $2 \leq i \leq n$, we denote $[L_t]_{i,i-1}$ by α_t^{i-1}, so that $\alpha_t^{i-1} \in \mathcal{M}_d(\mathbb{R})$.We may summarize the form of $\mathbf{L}_t$ as follows: for $\mathbf{x} = (x_1, \dots, x_n) \in (\mathbb{R}^d)^n$, $\mathbf{L}_t \mathbf{x} = (0, \alpha_t^1 x_1, \dots, \alpha_t^{n-1} x_{n-1})^* + \mathbf{U}_t \mathbf{x}$, where $\mathbf{U}_t \in \mathcal{M}_{nd}(\mathbb{R})$ is an "upper triangular" block matrix, i.e. the subdiagonal blocks of $\mathbf{U}_t$ are zero. (See the picture again.) (*In comparison with* ($\mathbf{A}$), ($\mathbf{A}^{\text{linear}}$) *doesn't refer to the subsets* $\mathcal{E}_1, \dots, \mathcal{E}_{n-1}$. *We introduce them later.*)

For an initial condition $\mathbf{x} = (x_1, \dots, x_n) \in (\mathbb{R}^d)^n$ and a function φ in the space $L^2([0, T], \mathbb{R}^d)$, we consider the controlled differential system

$$\dot{\mathbf{S}}_t = \mathbf{L}_t \mathbf{S}_t + B \Sigma_t \varphi_t, \quad a.e.\ t \in [0, T] \quad ; \quad \mathbf{S}_0 = \mathbf{x}, \tag{3.1}$$

whose solution is denoted by $(\mathbf{S}_t(\varphi))_{0 \leq t \leq T}$. Equation (3.1) may be understood as a very simple modelling for the propagation of a forcing φ into a chain of oscillators, the impulse in the i^{th} oscillator being transmitted from the $(i-1)^{\text{th}}$ oscillator by the matrix α^{i-1}.

In what follows, we plug the derivative of a Brownian path as an entry of the above system: for an $\mathbb{R}^d$ valued Brownian motion $(W_t)_{t \geq 0}$, we consider the random process $(\mathbf{G}_t = \mathbf{S}_t(\dot{W}))_{0 \leq t \leq T}$. It satisfies in the Itô sense the stochastic version of (3.1): $d\mathbf{G}_t = \mathbf{L}_t \mathbf{G}_t dt + B \Sigma_t dW_t$, $t \in [0, T]$, which is a particular example of (1.1). Obviously, $(\mathbf{G}_t)_{0 \leq t \leq T}$ is a Gaussian process. For each $t \in [0, T]$, we denote by $\mathbf{K}_t$ the covariance matrix of the Gaussian vector $\mathbf{G}_t$. Here is, in the current specific Gaussian framework, the connection between the absolute continuity of the law of $\mathbf{G}_T$ and the deterministic control problem (2.8)-(2.9) when (2.8) has the dynamics (3.1):

Proposition 3.1 *For an arbitrary $T > 0$, let ($\mathbf{A}^{\text{linear}}$) be in force. Then, the matrix $\mathbf{K}_T$ is non-degenerate if and only if the deterministic system* (3.1) *is controllable, i.e. if for any initial condition $\mathbf{x} \in \mathbb{R}^{nd}$ and terminal point $\mathbf{y} \in \mathbb{R}^{nd}$, there exists φ in $L^2([0, T], \mathbb{R}^d)$ such that $\mathbf{S}_0(\varphi) = \mathbf{x}$ and $\mathbf{S}_T(\varphi) = \mathbf{y}$. In such a case, the density of $\mathbf{G}_T$, with initial condition $\mathbf{G}_0 = \mathbf{x}$ and terminal condition $\mathbf{G}_T = \mathbf{y}$, has the form*

$$q(0, T, \mathbf{x}, \mathbf{y}) = (2\pi)^{-nd/2} \det{}^{-1/2}(\mathbf{K}_T) \exp\Big(-\frac{I_{\text{linear}}(T, \mathbf{x}, \mathbf{y})}{2}\Big), \tag{3.2}$$

[6] $\mathbf{L}_t$ has the form $\begin{pmatrix} \alpha_t & U_t \\ 0 & \end{pmatrix}$

where $I_{\mathrm{linear}}(T,\mathbf{x},\mathbf{y})$ *is the minimal cost*

$$I_{\mathrm{linear}}(T,\mathbf{x},\mathbf{y}) = \inf\Big\{\int_0^T |\varphi_t|^2 dt : \mathbf{S}_0(\varphi)=\mathbf{x},\ \mathbf{S}_T(\varphi)=\mathbf{y}\Big\}, \tag{3.3}$$

and is equal to

$$I_{\mathrm{linear}}(T,\mathbf{x},\mathbf{y}) = \langle \mathbf{R}(T,0)\mathbf{x}-\mathbf{y}, \mathbf{K}_T^{-1}[\mathbf{R}(T,0)\mathbf{x}-\mathbf{y}]\rangle, \tag{3.4}$$

where $(\mathbf{R}(t,t_0))_{0\le t_0,t\le T}$ *stands for the resolvent associated with* $(\mathbf{L}_t)_{0\le t\le T}$*, i.e.* $[d/dt][\mathbf{R}(t,t_0)] = \mathbf{L}_t[\mathbf{R}(t,t_0)]$ *and* $\mathbf{R}(t_0,t_0) = I_{nd}$*, identity matrix of size* $nd\times nd$*. A sufficient condition for controllability, or equivalently to guarantee* $\det(\mathbf{K}_T)>0$*, is:*

$$\text{for a.e. } t\in[0,T] \text{ and } i\in\{1,\dots,d\},\ \det(\alpha_t^i)>0. \tag{3.5}$$

(In control theory, Eq. (3.4) *says that* $\mathbf{K}_T$ *is the Gram matrix associated with* I_{linear}*.)*

Proof. By Coron [Cor07, Theorem 1.11, Chap. 1], the controllability problem associated with the controlled equation (3.1) admits the matrix $\mathbf{Q}_T = \int_0^T \mathbf{R}(T,t)BA_tB^*[\mathbf{R}(T,t)]^*dt$ as Gram matrix at time T. In particular, the problem (3.1) is controllable at time T if and only $\mathbf{Q}_T$ is invertible. In such a case, the cost $I_{\mathrm{linear}}(T,\mathbf{x},\mathbf{y})$ as defined in (3.3) is given by

$$I_{\mathrm{linear}}(T,\mathbf{x},\mathbf{y}) = \langle \mathbf{R}(T,0)\mathbf{x}-\mathbf{y}, (\mathbf{Q}_T)^{-1}[\mathbf{R}(T,0)\mathbf{x}-\mathbf{y}]\rangle, \tag{3.6}$$

see Proposition 1.13 in the same reference.

We now prove that $\mathbf{K}_T=\mathbf{Q}_T$. (By (3.6), this will prove (3.4).) Indeed, $d\mathbf{G}_t = \mathbf{L}_t\mathbf{G}_tdt + B\Sigma_tdW_t$, $0\le t\le T$. Using the resolvent, it may be written $\mathbf{G}_t = \mathbf{R}(t,0)\mathbf{x} + \int_0^t \mathbf{R}(t,s)B\Sigma_sdW_s$, so that

$$\begin{aligned}\mathbf{K}_T &= \mathbb{E}\Big[\int_0^T \mathbf{R}(T,s)B\Sigma_sdW_s\Big(\int_0^T \mathbf{R}(T,s)B\Sigma_sdW_s\Big)^*\Big]\\ &= \int_0^T \mathbf{R}(T,s)BA_sB^*[\mathbf{R}(T,s)]^*ds = \mathbf{Q}_T.\end{aligned} \tag{3.7}$$

This proves that $\mathbf{G}_T$, with the initial condition $\mathbf{G}_0=\mathbf{x}$, admits a density if and only if the problem (3.1) is controllable. Moreover, $\mathbf{G}_T$ admits $\mathbf{Q}_T$ as covariance matrix and $\mathbf{R}(T,0)\mathbf{x}$ as mean under the initial condition $\mathbf{G}_0$: by (3.4), we obtain (3.2).

It finally remains to check that (3.5) implies the controllability of (3.1). By Sontag [Son98, Lemma 3.5.8], it is controllable if and only if any solution $\boldsymbol{\phi}:[0,T]\mapsto\mathbb{R}^{nd}$ to the differential system

$$\dot{\boldsymbol{\phi}}_t = -\mathbf{L}_t^*\boldsymbol{\phi}_t,\ \Sigma_t^*B^*\boldsymbol{\phi}_t = 0 \quad a.e.\ t\in[0,T], \tag{3.8}$$

is zero a.e. on $[0,T]$. To complete the proof, we just have to check that the above sufficient and necessary condition holds under (3.5). For $\boldsymbol{\phi}$ fulfilling (3.8), we write $\boldsymbol{\phi}_t = (\phi_t^1, \dots, \phi_t^n)$, each block being of size d. Since Σ is non-degenerate, the condition $\Sigma_t^* B^* \boldsymbol{\phi}_t = 0$ in (3.8) says that $\phi^1 = 0$ on $[0,T]$. By the "subdiagonal+upper triangular" form of $\mathbf{L}_t$, the first line of the equation $\dot{\boldsymbol{\phi}}_t = -\mathbf{L}_t^* \boldsymbol{\phi}_t$ says that $\dot{\phi}_t^1 = c_t \phi_t^1 - \alpha_t^1 \phi_t^2$, for some bounded measurable coefficient $(c_t)_{0 \le t \le T}$. Since ϕ^1 is zero, we deduce that $\phi^2 = 0$ a.e. on $[0,T]$ if the matrix α_t^1 is invertible for a.e. $t \in [0,T]$. By induction, we deduce that $\boldsymbol{\phi}$ is zero if, for a.e. $t \in [0,T]$, each matrix α_t^i, $1 \le i \le n-1$, is invertible. □

For our own purpose, we are interested in Gaussian processes $(\mathbf{G}_t)_{0 \le t \le T}$ with the right time scale, given by the scale matrices $(\mathbb{T}_t)_{0 < t \le T}$ defined in Theorem 1.1:

Definition 3.2 *We say that the Gaussian process $(\mathbf{G}_t)_{0 \le t \le T}$ satisfies a good scaling property of parameter $c \ge 1$ if, for all $t \in (0,T]$, the covariance matrix $\mathbf{K}_t$ satisfies $c^{-1} t^{-1} |\mathbb{T}_t \mathbf{x}|^2 \le \langle \mathbf{K}_t \mathbf{x}, \mathbf{x} \rangle \le c t^{-1} |\mathbb{T}_t \mathbf{x}|^2$ for any $\mathbf{x} \in \mathbb{R}^{nd}$. (That is, the spectrum of $t \mathbb{T}_t^{-1} \mathbf{K}_t \mathbb{T}_t^{-1}$, or equivalently of $t^{-1} \mathbb{T}_t \mathbf{K}_t^{-1} \mathbb{T}_t$, belongs to $[c^{-1}, c]$ for any $t \in (0,T]$. In particular, $c^{-1} t |\mathbb{T}_t^{-1} \mathbf{y}|^2 \le \langle \mathbf{y}, \mathbf{K}_t^{-1} \mathbf{y} \rangle \le c t |\mathbb{T}_t^{-1} \mathbf{y}|^2$ for $t \in (0,T]$.)*

The above definition may be interpreted as follows: under the good scaling property of parameter c, $\mathbf{G}_t$ has the canonical writing $\mathbf{G}_t = t^{-1/2} \mathbb{T}_t \hat{\mathbf{G}}_1$, the spectrum of the covariance matrix of the Gaussian vector $\hat{\mathbf{G}}_1$ being between c^{-1} and c. In other words, the i^{th} coordinate G^i of the process $\mathbf{G}$ lives at scale $t^{i-1/2}$, $1 \le i \le n$. Moreover, $c^{-nd} t^{n^2 d} \le \det(\mathbf{K}_t) \le c^{nd} t^{n^2 d}$ for $t \in (0,T]$: this coincides with the "diagonal" decay in Theorem 1.1.

By Proposition 3.1 and Definition 3.2, we deduce

Proposition 3.3 *Let $(\mathbf{A}^{\text{linear}})$ hold. Assume also that the Gaussian process $(\mathbf{G}_t)_{0 \le t \le T}$ admits a good scaling property of parameter $c \ge 1$, then there exists a constant $C_{3.3}(c)$, only depending on c, κ, Λ, d and n, such that, for any $t \in (0,T]$ and any starting point $\mathbf{x} \in \mathbb{R}^{nd}$, the law of $\mathbf{G}_t$, with $\mathbf{G}_0 = \mathbf{x}$, admits a density $(q(t,\mathbf{x},\mathbf{y}))_{\mathbf{y} \in \mathbb{R}^{nd}}$ satisfying*

$$\begin{aligned} & C_{3.3}^{-1} t^{-n^2 d/2} \exp\Big(-C_{3.3} t \Big| \mathbb{T}_t^{-1} \Big[\mathbf{R}(t,0)\mathbf{x} - \mathbf{y} \Big] \Big|^2 \Big) \\ & \quad \le q(t,\mathbf{x},\mathbf{y}) \\ & \quad \le C_{3.3} t^{-n^2 d/2} \exp\Big(-C_{3.3}^{-1} t \Big| \mathbb{T}_t^{-1} \Big[\mathbf{R}(t,0)\mathbf{x} - \mathbf{y} \Big] \Big|^2 \Big). \end{aligned} \tag{3.9}$$

We emphasize that Proposition 3.3 is a preliminary version of Theorem 1.1 for linear systems satisfying the "good scaling property". The next proposition provides the complete version of Theorem 1.1 for linear systems satisfying $(\mathbf{A})$. It will play a key role when investigating the linearized version of (1.1) (see

Sections 2 and 4).

Proposition 3.4 *Assume that, in addition to* ($\mathbf{A}^{\text{linear}}$)*, for all* $i \in \{1, \dots, n-1\}$ *and* $t \in [0,T]$*,* α_t^i *belongs to* $\mathcal{E}_i$ *(closed convex subset of* $GL_d(\mathbb{R})$*). Then, there exists a constant* $c \geq 1$*, only depending on* $\mathcal{E}_1, \dots, \mathcal{E}_{n-1}$*, on* κ*,* Λ*,* d*,* n *and* T*, such that* $(\mathbf{G}_t)_{0 \leq t \leq T}$ *satisfies a good scaling property of parameter* c*. In particular, Proposition 3.3 applies.*

Before we prove Proposition 3.4, we first notice that it is not sufficient to have a lower bound on the determinants of $\alpha^1, \dots, \alpha^{n-1}$ (or equivalently on the spectrum of $\alpha^1(\alpha^1)^*, \dots, \alpha^{n-1}(\alpha^{n-1})^*$) to control from above and from below the spectrum of the covariance matrix $\mathbf{K}_T$. To do so, an additional assumption is needed on the non-degeneracy of the matrices $\alpha^1, \dots, \alpha^{n-1}$. The following example explains the role of the closed convex subsets $\mathcal{E}_1, \dots, \mathcal{E}_{n-1}$ that appear in Assumption ($\mathbf{A}$) and Proposition 3.4.

Example 3.5 *Set* $\beta_t^m = \begin{pmatrix} \cos(2\pi m t) & -\sin(2\pi m t) \\ \sin(2\pi m t) & \cos(2\pi m t) \end{pmatrix}$ *for* $m \geq 1$ *and* $t \in [0,1]$*, so that* $\det(\beta_t^m) = 1$ *for all* $m \geq 1$ *and* $t \in [0,1]$*. Then, for* $T = 1$*,* $d = n = 2$ *(so that* $(W_t)_{t \geq 0}$ *is a 2-dimensional Brownian motion) and* $\mathbf{L}_t = \begin{pmatrix} 0 & 0 \\ \beta_t^m & 0 \end{pmatrix}$*,* $0 \leq t \leq 1$*, the process* $(\mathbf{G}_t = (G_t^1, G_t^2))_{0 \leq t \leq 1}$ *satisfies* $dG_t^1 = dW_t$ *and* $dG_t^2 = \beta_t^m G_t^1 dt$*. In particular, for* $\mathbf{G}_0 = 0$*,* $G_1^2 = \int_0^1 \beta_t^m W_t dt$*, so that the first coordinate of* G_1^2 *is equal to* $Z = \int_0^1 \cos(2\pi m t) W_t^1 dt - \int_0^1 \sin(2\pi m t) W_t^2 dt$*. We let the reader check that* $\mathbb{V}(Z) = O(m^{-2})$ *as* m *tends to* $+\infty$*. This shows that* $\det(K_1)$ *vanishes with* $m \to +\infty$ *whereas* $\det(\beta_t^m)$ *is constant.*

As shown in the proof of Proposition 3.4, the dramatic point in the above example is that the sequence of functions $(t \in [0,1] \mapsto \beta_t^m)_{m \geq 0}$ weakly converges towards 0 in $L^2([0,1], \mathcal{M}_2(\mathbb{R}))$. This is an obvious consequence of the Riemann-Lebesgue theorem.

Proof (Proposition 3.4). Let us first suppose $T = 1$. By the representation formula (3.7), we can also assume w.l.o.g. that A_t is the identity matrix for any $t \in [0,1]$. Up to a rotation of W, this is equivalent to assume that Σ is the identity matrix for any $t \in [0,1]$. We then define $\boldsymbol{\mathcal{E}}$ as the set of mappings $(t \in [0,1] \mapsto \mathbf{L}_t \in \mathcal{M}_{nd}(\mathbb{R}))$, bounded by κ for a.e. $t \in [0,1]$, such that the $d \times d$ block $[L_t]_{i,j}$, $1 \leq i,j \leq n$, is zero if $i \geq j+2$ and belongs to $\mathcal{E}^{i-1}$ if $j = i-1$, for a.e. $t \in [0,1]$.

It is well seen that, for each $t_0 \in [0,1]$, the mapping $\mathbf{L} \in L^2([0,1], \mathcal{M}_{nd}(\mathbb{R}) \mapsto \mathbf{R}(t_0, \cdot) \in \mathcal{C}([0,1], \mathcal{M}_{nd}(\mathbb{R}))$ is continuous, $L^2([0,1], \mathcal{M}_{nd}(\mathbb{R}))$ being equipped with the weak topology and $\mathcal{C}([0,1], \mathcal{M}_{nd}(\mathbb{R}))$ with the uniform convergence topology. Using the representation formula (3.7), it is then clear that the mapping $\mathbf{L} \in L^2([0,1],\mathcal{M}_{nd}(\mathbb{R})) \mapsto \mathbf{K}_1 \in \mathcal{M}_{nd}(\mathbb{R})$ is continuous, $L^2([0,1],\mathcal{M}_{nd}(\mathbb{R}))$ being equipped with the weak topology. By Proposition 3.1, we know that $\det(\mathbf{K}_1) > 0$ for any $\mathbf{L} \in \boldsymbol{\mathcal{E}}$. If $\boldsymbol{\mathcal{E}}$ is a compact subset of $L^2([0,1],\mathcal{M}_{nd}(\mathbb{R}))$ equipped with the weak topology, then $\exists \gamma \in (0,1]$ s.t. $\inf\{\det(\mathbf{K}_1),\ \mathbf{L} \in \boldsymbol{\mathcal{E}}\} \geq \gamma$ and $\sup\{\|\mathbf{K}_1\|,\ \mathbf{L} \in \boldsymbol{\mathcal{E}}\} \leq \gamma^{-1}$, i.e. the spectrum of $\mathbf{K}_1$ is bounded from above and from below, uniformly in $\mathbf{L} \in \boldsymbol{\mathcal{E}}$: $\exists c \geq 1$ as in the statement of Proposition 3.4 such that $c^{-1}|\mathbf{x}|^2 \leq \langle \mathbf{K}_1 \mathbf{x}, \mathbf{x}\rangle \leq c|\mathbf{x}|^2$ for all $\mathbf{x} \in \mathbb{R}^{nd}$. It thus remains to prove that $\boldsymbol{\mathcal{E}}$ is a compact subset of $L^2([0,1],\mathcal{M}_{nd}(\mathbb{R}))$ equipped with the weak topology: since it is bounded, it is sufficient to prove that it is closed for the weak topology. Because of the convexity of $\mathcal{E}_1,\dots,\mathcal{E}_{n-1}$, it is convex: since it is clearly closed for the strong topology on $L^2([0,1],\mathcal{M}_{nd}(\mathbb{R}))$, it is also closed for the weak topology.

To investigate $\mathbf{K}_t$, $t \in (0,1)$, we use a linear variant of the scaling property (2.11). For t fixed in $(0,1)$, we can define similarly to Section 2.3: $\hat{\mathbf{G}}_s^t := t^{1/2}\mathbb{T}_t^{-1}\mathbf{G}_{st}$, $s \in [0,1]$. By the scaling Lemma 3.6 below (see (3.10)), the process $(\hat{\mathbf{G}}_s^t)_{0\leq s\leq 1}$ satisfies the same properties as $(\mathbf{G}_s)_{0\leq s\leq 1}$ so that the covariance matrix $\hat{\mathbf{K}}_1^t := \mathrm{Cov}(\hat{\mathbf{G}}_1^t)$ satisfies: $c^{-1}|\mathbf{x}|^2 \leq \langle \mathbf{x}, \hat{\mathbf{K}}_1^t \mathbf{x}\rangle \leq c|\mathbf{x}|^2$ for all $\mathbf{x} \in \mathbb{R}^{nd}$ and for the same c as above. Again by Lemma 3.6 below, $\mathbf{K}_t = t^{-1}\mathbb{T}_t\hat{\mathbf{K}}_1^t T_t$. The good scaling property easily follows.

The case $T \neq 1$ can be handled, up to a suitable modification of the Lipschitz constants for $T > 1$, by similar scaling arguments. $\square$

Here is the scaling lemma. (It is nothing but a linear version of (2.10)–(2.11)–(2.12).)

Lemma 3.6 (Scaling Lemma) *Let $T > 0$ and $(\mathbf{L}_s, \Sigma_s)_{s\in[0,T]}$ be as in Proposition 3.4 (i.e. $(\mathbf{A}^{\text{linear}})$ holds and, for all $i \in \{1,\dots,n-1\}$ and $s \in [0,T]$, α_s^i belongs to $\mathcal{E}_i$). Fix $t \in (0,T]$ and set $\hat{\mathbf{G}}_s^t = t^{1/2}\mathbb{T}_t^{-1}\mathbf{G}_{st}$, $0 \leq s \leq 1$. Then, $(\hat{\mathbf{G}}_s^t)_{0\leq s\leq 1}$ satisfies (3.1) with respect to $(\hat{\mathbf{L}}_s^t = t\mathbb{T}_t^{-1}\mathbf{L}_{st}\mathbb{T}_t)_{0\leq s\leq 1}$, $(\hat{\Sigma}_s^t = \Sigma_{ts})_{0\leq s\leq 1}$ and $(\hat{W}_s^t = t^{-1/2}W_{st})_{0\leq s\leq 1}$. The resolvent $[\hat{\mathbf{R}}^t(s_1,s_0)]_{0\leq s_0,s_1\leq 1}$ associated with $(\hat{\mathbf{L}}_s^t)_{0\leq s\leq 1}$ has the form $\hat{\mathbf{R}}^t(s_1,s_0) = \mathbb{T}_t^{-1}\mathbf{R}(s_1 t, s_0 t)\mathbb{T}_t$, $s_0, s_1 \in [0,1]$ and the covariance matrix of $\hat{\mathbf{G}}_s^t$, $0 \leq s \leq 1$, is given by $\hat{\mathbf{K}}_s^t := \mathrm{Cov}(\hat{\mathbf{G}}_s^t) = t\mathbb{T}_t^{-1}\mathbf{K}_{st}\mathbb{T}_t^{-1}$. The matrices $(\hat{\mathbf{L}}_s^t)_{0\leq s\leq 1}$ and $(\hat{\Sigma}_s^t)_{0\leq s\leq 1}$ satisfy for all $0 \leq s \leq 1$:*

$$
\begin{aligned}
&|\hat{\mathbf{L}}_s^t| \leq (1 \vee T^n)\kappa, \\
&[\hat{\mathbf{L}}_s^t]_{i,i-1} \in \mathcal{E}_{i-1},\ 2 \leq i \leq n,\ ;\ [\hat{\mathbf{L}}_s^t]_{i,j} = 0,\ 1 \leq j \leq j+2 \leq i \leq n, \\
&\mathrm{Spectrum}(\hat{\Sigma}_s^t(\hat{\Sigma}_s^t)^*) \subset [\Lambda^{-1}, \Lambda].
\end{aligned}
\tag{3.10}
$$

In particular, for $t \leq T \leq 1$, $(\hat{\mathbf{L}}_s^t)_{0\leq s\leq 1}$ and $(\hat{\Sigma}_s^t)_{0\leq s\leq 1}$ satisfy $(\mathbf{A}^{\text{linear}})$ independently of t and T: by Proposition 3.4, $(\hat{\mathbf{K}}_s^t)_{0\leq s\leq 1}$ satisfy the good scaling property with respect to some c depending on $(\mathbf{A})$ only. Similarly, the resolvent $(\hat{\mathbf{R}}^t(s_1,s_0))_{0\leq s_0,s_1\leq 1}$ can be bounded independently of t and T.

Proof. Following (2.10), it is easily checked that $d\hat{\mathbf{G}}_s^t = \hat{\mathbf{L}}_s^t \hat{\mathbf{G}}_s^t ds + B\hat{\Sigma}_s^t d\hat{W}_s$, $0 \le s \le 1$. Similarly, for a mapping $\boldsymbol{\phi}_{0\le s\le T}$ satisfying $[d/ds]\boldsymbol{\phi}_s = \mathbf{L}_s\boldsymbol{\phi}_s$, $0 \le s \le T$, $(\hat{\boldsymbol{\phi}}_s^t := \mathbb{T}_t^{-1}\boldsymbol{\phi}_{st})_{0\le s\le 1}$ satisfies $[d/ds][\hat{\boldsymbol{\phi}}_s^t] = \hat{\mathbf{L}}_s^t \hat{\boldsymbol{\phi}}_s^t$. In particular, recalling that $\hat{\mathbf{R}}$ denotes the resolvent associated with $\hat{\mathbf{L}}^t$, $\hat{\mathbf{R}}^t(s_1, s_0) = \mathbb{T}_t^{-1}\mathbf{R}(s_1 t, s_0 t)\mathbb{T}_t$. We also let the reader check from (3.7) that $\hat{\mathbf{K}}_s^t = t\mathbb{T}_t^{-1}\mathbf{K}_{st}\mathbb{T}_t^{-1}$.

Finally, we note that $[\hat{\mathbf{L}}_s^t]_{i,j} = t^{j+1-i}[\mathbf{L}_s]_{i,j}$. Since $[\mathbf{L}_s]_{i,j} = 0$ for $1 \le j \le j+2 \le i \le n$, we deduce that $|[\hat{\mathbf{L}}_s^t]_{i,j}| \le (1 \vee T^n)|[\mathbf{L}_s]_{i,j}|$. We also have $[\hat{\mathbf{L}}_s^t]_{i,j} = 0$ for $1 \le j \le j+2 \le i \le n$ and $[\hat{\mathbf{L}}_s^t]_{i,i-1} = [\mathbf{L}_s]_{i,i-1} \in \mathcal{E}^{i-1}$ for $2 \le i \le n$. The bounds for the spectrum of $\hat{\Sigma}_s^t(\hat{\Sigma}_s^t)^*$ are obvious. □

3.2 Stochastic Control Problem and Representation of the Gaussian Density

As announced in Point (i) of Subsection 2.2, we now specify the stochastic control formulation (2.6)-(2.7) in the Gaussian case and investigate the form and the properties of the optimal control. To simplify, we only consider the case when the initial and final points in (2.6) are $\mathbf{0}$: as explained in Point (iii), this is the only case needed for the main proof of Theorem 1.1. (Actually, the analysis could be extended to arbitrary initial and final points without additional difficulty.)

Proposition 3.7 *Let $T > 0$ and ($\mathbf{A}^{\mathrm{linear}}$) hold. Assume in addition that, for all $i \in \{1, \ldots, n-1\}$ and $t \in [0, T]$, α_t^i belongs to $\mathcal{E}_i$. Then, $(\mathbf{G}_t)_{0\le t\le T}$ admits a transition kernel $(q(s,t,\mathbf{x},\mathbf{y}))_{0\le s<t\le T;\mathbf{x},\mathbf{y}\in\mathbb{R}^{nd}}$ and for $\mathbf{G}_0 = \mathbf{0}$, the density $q(0, T, \mathbf{0}, \mathbf{0})$ of $\mathbf{G}_T$ at $\mathbf{0}$ admits the representation: for all $\varepsilon \in (0, T]$,*

$$
\begin{aligned}
-\ln(q(0,T,\mathbf{0},\mathbf{0})) = -\ln(q(T-\varepsilon,T,\boldsymbol{\Gamma}_{T-\varepsilon},\mathbf{0})) &+ \frac{1}{2}\int_0^{T-\varepsilon} \langle A_t^{-1}\gamma_t, \gamma_t\rangle dt \\
&+ \int_0^{T-\varepsilon} \langle \Sigma_t^{-1}\gamma_t, dW_t\rangle,
\end{aligned}
\tag{3.11}
$$

where $(\boldsymbol{\Gamma}_t)_{0\le t<T}$ is the solution of the controlled SDE

$$
d\boldsymbol{\Gamma}_t = \mathbf{L}_t\boldsymbol{\Gamma}_t dt + B\gamma_t dt + B\Sigma_t dW_t, \ 0 \le t < T, \tag{3.12}
$$

with $\boldsymbol{\Gamma}_0 = \mathbf{0}$ and, for all $0 \le t < T$,

$$
\begin{aligned}
&\gamma_t = -A_t B^*[\mathbf{R}(T,t)]^* \int_0^t \boldsymbol{\rho}_s^{-1} B\Sigma_s dW_s, \\
&\boldsymbol{\rho}_t = \mathbf{H}(t,T)[\mathbf{R}(T,t)]^*, \ \mathbf{H}(t,T) = \int_t^T \mathbf{R}(t,s)BA_sB^*[\mathbf{R}(t,s)]^* ds.
\end{aligned}
\tag{3.13}
$$

(The matrices $\boldsymbol{\rho}_t$ and $\mathbf{H}(t,T)$ are of size $nd \times nd$.) Moreover, for all $t \in [0, T)$, $\boldsymbol{\rho}_t$ is invertible and, for all $\varepsilon \in (0, T]$, there exists a constant $C_{3.7}(\varepsilon) > 0$, only

depending on ε, T and $(\mathbf{A})$, such that

$$|\boldsymbol{\rho}_t^{-1}| \leq C_{3.7}(\varepsilon), \quad 0 \leq t \leq T - \varepsilon. \tag{3.14}$$

(In particular, for all $\varepsilon \in (0,T]$, $\mathbb{E}\int_0^{T-\varepsilon} |\gamma_t|^2 dt$ is finite.) Finally, $\boldsymbol{\Gamma}_t$ admits the representation

$$\boldsymbol{\Gamma}_t = \boldsymbol{\rho}_t \int_0^t \boldsymbol{\rho}_s^{-1} B\Sigma_s dW_s. \tag{3.15}$$

Proof. Existence of the transition densities $(q(s,t,\mathbf{x},\mathbf{y}))_{0\leq s<t\leq T;\mathbf{x},\mathbf{y}\in\mathbb{R}^{nd}}$ follows from Propositions 3.4 and 3.3: any vector $(\mathbf{G}_{s+h} - \mathbf{G}_s)_{0\leq s<s+h\leq T}$ fulfills the required assumptions in these two propositions. By (3.7) and the semi-group property for the resolvent $\mathbf{R}$ (i.e. $\mathbf{R}(T,t)\mathbf{R}(t,s) = \mathbf{R}(T,s)$), we also emphasize that

$$\mathbf{R}(T,t)\mathbf{H}(t,T)[\mathbf{R}(T,t)]^* = \int_t^T \mathbf{R}(T,s)BA_sB^*[\mathbf{R}(T,s)]^* ds$$

is the covariance matrix of $\mathbf{G}_T - \mathbf{G}_t$. We denote it by $\mathbf{K}(t,T)$: by the scaling Lemma 3.6, we can write $\mathbf{K}(t,T) = (T-t)^{-1}\mathbb{T}_{(T-t)}\hat{\mathbf{K}}_1^{t,T}\mathbb{T}_{(T-t)}$ and $\mathbf{R}(t,T) = \mathbb{T}_{T-t}\hat{\mathbf{R}}^{t,T}(0,1)\mathbb{T}_{T-t}^{-1}$, $\hat{\mathbf{K}}_1^{t,T}$, $\hat{\mathbf{R}}^{t,T}(0,1)$ standing for the covariance and resolvent matrices at time 1 associated with the rescaled process $\hat{\mathbf{G}}_u^{t,T} := (T-t)^{1/2}\mathbb{T}_{T-t}^{-1}$ $(\mathbf{G}_{t+u(T-t)} - \mathbf{G}_t)$, $u \in [0,1]$, satisfying the same assumptions as $\mathbf{G}$ uniformly in the scaling parameter $t \in [0,T)$. Therefore, we have $\mathbf{H}(t,T) = \mathbf{R}(t,T)\mathbf{K}(t,T)[\mathbf{R}(t,T)]^* = (T-t)^{-1}\mathbb{T}_{T-t}\hat{\mathbf{R}}^{t,T}(0,1)\hat{\mathbf{K}}_1^{t,T}[\hat{\mathbf{R}}^{t,T}(0,1)]^*\mathbb{T}_{T-t}$. Since the rescaled process satisfies the same assumptions as $\mathbf{G}$ (the constant κ being possibly magnified by some power of T, see (3.10)), we know that $\hat{\mathbf{K}}_1^{t,T}$ is non-degenerate (uniformly in t), so that the spectrum of $\mathbf{H}(t,T)$ is bounded from below by a positive constant, uniformly in $t \in [0, T-\varepsilon)$, for any $\varepsilon \in (0,T]$. Inequality (3.14) easily follows. This shows that, for any $\varepsilon \in (0,T]$, $\mathbb{E}\int_0^{T-\varepsilon} |\gamma_t|^2 dt < +\infty$. As a consequence, Eq. (3.12) is well-posed.

We now provide the explicit form of $(\boldsymbol{\Gamma}_t)_{0\leq t<T}$ by variation of parameters. To do so, we write the ODE satisfied by the matrices $(\mathbf{H}(t,T))_{0\leq t\leq T}$. It satisfies

$$[d/dt]\mathbf{H}(t,T) = \mathbf{L}_t\mathbf{H}(t,T) + \mathbf{H}(t,T)\mathbf{L}_t^* - BA_tB^*, \quad 0 \leq t \leq T. \tag{3.16}$$

Since $\boldsymbol{\rho}_t = \mathbf{H}(t,T)[\mathbf{R}(T,t)]^*$ for all $t \in [0,T]$, we have

$$\dot{\boldsymbol{\rho}}_t = \mathbf{L}_t\boldsymbol{\rho}_t - BA_tB^*[\mathbf{R}(T,t)]^* \tag{3.17}$$

since $[d/dt][\mathbf{R}(T,t)]^* = -\mathbf{L}_t^*[\mathbf{R}(T,t)]^*$. (This is an easy consequence of the inverse property $\mathbf{R}(T,t) = [\mathbf{R}(t,T)]^{-1}$.) We then see $(\boldsymbol{\rho}_t)_{0\leq t\leq T}$ as a kind of resolvent vanishing at time T. By (3.14), $(\tilde{\boldsymbol{\Gamma}}_t = \boldsymbol{\rho}_t \int_0^t \boldsymbol{\rho}_s^{-1} B\Sigma_s dW_s)_{0\leq t<T}$ is well defined: it satisfies $d\tilde{\boldsymbol{\Gamma}}_t = \mathbf{L}_t\tilde{\boldsymbol{\Gamma}}_t dt + B\gamma_t dt + B\Sigma_t dW_t$. Comparing with (3.12), we deduce that $\boldsymbol{\Gamma}_t = \tilde{\boldsymbol{\Gamma}}_t$ for all $t \in [0,T)$ a.s. This proves (3.15).

We finally set $J_t := -2\ln[q(t,T,\boldsymbol{\Gamma}_t,\mathbf{0})]$, $0 \leq t < T$, that is $J_t = nd\ln(2\pi) + \ln(\det(\mathbf{K}(t,T))) + \langle[\mathbf{K}(t,T)]^{-1}\mathbf{R}(T,t)\boldsymbol{\Gamma}_t, \mathbf{R}(T,t)\boldsymbol{\Gamma}_t\rangle$. (See Proposition 3.1 for the form of $q(t,T,\mathbf{\Gamma}_t,\mathbf{0})$.) From the equality $\mathbf{K}(t,T)=\mathbf{R}(T,t)\mathbf{H}(t,T)[\mathbf{R}(T,t)]^*$ and the Wronskian identity, we obtain

$$\det(\mathbf{K}(t,T))=\det(\mathbf{H}(t,T))\det(\mathbf{R}(T,t))^2=\det(\mathbf{H}(t,T))\exp(2\int_t^T \mathrm{Tr}(\mathbf{L}_u)du).$$

Hence, $J_t=nd\ln(2\pi)+\ln(\det(\mathbf{H}(t,T)))+2\int_t^T \mathrm{Tr}(\mathbf{L}_u)du+\langle[\mathbf{H}(t,T)]^{-1}\mathbf{\Gamma}_t,\mathbf{\Gamma}_t\rangle$. We then compute (use that $[d/dt](\ln(\det(\mathbf{f}_t))=\mathrm{Tr}(\mathbf{f}_t^{-1}[d\mathbf{f}_t/dt])$ for an $\mathcal{M}_{nd}(\mathbb{R})$-valued differentiable function $(\mathbf{f}_t)_{0\le t\le T}$ of non-zero determinant)

$$\begin{aligned}dJ_t&=\mathrm{Tr}\Big[[\mathbf{H}(t,T)]^{-1}[d/dt][\mathbf{H}(t,T)\Big]dt-2\mathrm{Tr}(\mathbf{L}_t)dt\\&\quad-\langle[d/dt][\mathbf{H}(t,T)][\mathbf{H}(t,T)]^{-1}\mathbf{\Gamma}_t,[\mathbf{H}(t,T)]^{-1}\mathbf{\Gamma}_t\rangle dt\\&\quad+2\langle[\mathbf{H}(t,T)]^{-1}\mathbf{\Gamma}_t,\mathbf{L}_t\mathbf{\Gamma}_t+B\gamma_t\rangle dt+\mathrm{Tr}\Big[BA_tB^*[\mathbf{H}(t,T)]^{-1}\Big]dt\\&\quad+2\langle[\mathbf{H}(t,T)]^{-1}\mathbf{\Gamma}_t,B\Sigma_t dW_t\rangle.\end{aligned}$$

By the ODE (3.16) satisfied by $(\mathbf{H}(t,T))_{0\le t\le T}$, the first line in the above right-hand side is equal to $-\mathrm{Tr}(BA_tB^*[\mathbf{H}(t,T)]^{-1})$ and the second line to $-2\langle[\mathbf{H}(t,T)]^{-1}\mathbf{\Gamma}_t,\mathbf{L}_t\mathbf{\Gamma}_t\rangle+\langle BA_tB^*[\mathbf{H}(t,T)]^{-1}\mathbf{\Gamma}_t,[\mathbf{H}(t,T)]^{-1}\mathbf{\Gamma}_t\rangle$, so that

$$\begin{aligned}dJ_t&=\Big(\langle BA_tB^*[\mathbf{H}(t,T)]^{-1}\mathbf{\Gamma}_t,[\mathbf{H}(t,T)]^{-1}\mathbf{\Gamma}_t\rangle+2\langle[\mathbf{H}(t,T)]^{-1}\mathbf{\Gamma}_t,B\gamma_t\rangle\Big)dt\\&\quad+2\langle[\mathbf{H}(t,T)]^{-1}\mathbf{\Gamma}_t,B\Sigma_t dW_t\rangle.\end{aligned}$$

By (3.13) and (3.15), $\gamma_t=-A_tB^*[\mathbf{R}(T,t)]^*\boldsymbol{\rho}_t^{-1}\mathbf{\Gamma}_t=-A_tB^*[\mathbf{H}(t,T)]^{-1}\mathbf{\Gamma}_t$, we deduce that $dJ_t=-\langle A_t^{-1}\gamma_t,\gamma_t\rangle dt-2\langle\Sigma_t^{-1}\gamma_t,dW_t\rangle$. □

Remark 3.8 *We mention that the result of Proposition 3.7 could also be derived from the formulation* (2.6) *by taking the density itself for the mollifier* η_ε, *i.e.* $\eta_\varepsilon(\mathbf{x})=q(T-\varepsilon,T,\mathbf{x},\mathbf{0})$, $\forall\mathbf{x}\in\mathbb{R}^{nd}$. *The optimal control is then given by* $(v_t^*)_{0\le t<T}$ *in* (2.7). *To derive the explicit formulas* (3.13) *and* (3.15), *this method actually leads to the same computations as the ones used above.*

Since $q(T-\varepsilon,T,\cdot,\mathbf{0})$ converges (in the weak sense) to the Dirac mass at $\mathbf{0}$ in (3.11) as ε tends to 0, we expect $\mathbf{\Gamma}_{T-\varepsilon}$ to converge to $\mathbf{0}$ with ε. The rate of convergence is given by

Lemma 3.9 *Under the assumption and notation of Proposition 3.7, for any* $p\ge 1$, *there exists a constant* $C_{3.9}(p)>0$, *only depending on* p, $T\vee 1$ *and* ($\mathbf{A}$), *such that* $\mathbb{E}[|(T-t)^{1/2}\mathbb{T}_{T-t}^{-1}\mathbf{\Gamma}_t|^p]\le C_{3.9}(p)$ *and* $\mathbb{E}[|\gamma_t|^p]\le C_{3.9}(p)(T-t)^{-p/2}$. *In particular, for all* $i\in\{1,\cdots,n\}$, $\mathbb{E}[|\mathbf{\Gamma}_t^i|]\le C(T-t)^{i-1/2}$.

Proof. For a given $t\in[0,T)$, we have $(T-t)^{1/2}\mathbb{T}_{T-t}^{-1}\mathbf{\Gamma}_t=(T-t)^{1/2}\mathbb{T}_{T-t}^{-1}\boldsymbol{\rho}_t\times\int_0^t\boldsymbol{\rho}_s^{-1}B\Sigma_s dW_s$. Obviously, it is a Gaussian random vector, with covariance matrix $(T-t)\mathbb{T}_{T-t}^{-1}\boldsymbol{\rho}_t[\int_0^t\boldsymbol{\rho}_s^{-1}BA_sB^*[\boldsymbol{\rho}_s^{-1}]^*ds]\boldsymbol{\rho}_t^*\mathbb{T}_{T-t}^{-1}$. By the scaling argument used in the proof of Proposition 3.7, we write

$$\begin{aligned}\boldsymbol{\rho}_t &= \mathbf{H}(t,T)[\mathbf{R}(T,t)]^* = \mathbf{H}(t,T)[(\mathbf{R}(t,T))^{-1}]^* \\ &= (T-t)^{-1}\mathbb{T}_{T-t}\hat{\mathbf{R}}^{t,T}(0,1)\hat{\mathbf{K}}_1^{t,T}\mathbb{T}_{T-t}.\end{aligned} \tag{3.18}$$

Denoting by $\leq$ the standard comparison between nonnegative symmetric matrices we obtain

$$\begin{aligned}&\int_0^t \boldsymbol{\rho}_s^{-1}BA_sB^*[\boldsymbol{\rho}_s^{-1}]^*ds \\ &\leq \Lambda \int_0^t (T-s)^2\mathbb{T}_{T-s}^{-1}[\hat{\mathbf{K}}_1^{s,T}]^{-1}\hat{\mathbf{R}}^{s,T}(1,0)\mathbb{T}_{T-s}^{-1}B \\ &\qquad\qquad \times B^*\mathbb{T}_{T-s}^{-1}[\hat{\mathbf{R}}^{s,T}(1,0)]^*[[\hat{\mathbf{K}}_1^{s,T}]^{-1}]^*\mathbb{T}_{T-s}^{-1}ds.\end{aligned}$$

We note that $\mathbb{T}_{T-s}^{-1}BB^*\mathbb{T}_{T-s}^{-1} = (T-s)^{-2}BB^*$. By Lemma 3.6, the rescaled matrices $[\hat{\mathbf{K}}_1^{s,T}]^{-1}$ and $\hat{\mathbf{R}}^{s,T}(1,0)$ are uniformly bounded in s by a constant only depending on $T \vee 1$ and $(\mathbf{A})$. As a consequence, there exists a nonnegative constant $C_{1\vee T}$ (which may vary from line to line), only depending on the parameters quoted in the statement, such that

$$\int_0^t \boldsymbol{\rho}_s^{-1}BA_sB^*[\boldsymbol{\rho}_s^{-1}]^*ds \leq C_{1\vee T}\int_0^t \mathbb{T}_{T-s}^{-2}ds \leq C_{1\vee T}(T-t)\mathbb{T}_{T-t}^{-2}. \tag{3.19}$$

Finally, the covariance matrix of $(T-t)^{1/2}\mathbb{T}_{T-t}^{-1}\boldsymbol{\Gamma}_t$ is bounded as follows:

$$\begin{aligned}&(T-t)\mathbb{T}_{T-t}^{-1}\boldsymbol{\rho}_t\left[\int_0^t \boldsymbol{\rho}_s^{-1}BA_sB^*[\boldsymbol{\rho}_s^{-1}]^*ds\right]\boldsymbol{\rho}_t^*\mathbb{T}_{T-t}^{-1} \\ &\leq C_{1\vee T}\hat{\mathbf{R}}^{t,T}(0,1)\hat{\mathbf{K}}_1^{t,T}\mathbb{T}_{T-t}\mathbb{T}_{T-t}^{-2}\mathbb{T}_{T-t}\hat{\mathbf{K}}_1^{t,T}[\hat{\mathbf{R}}^{t,T}(0,1)]^*.\end{aligned}$$

This clearly proves that the covariance matrix $(T-t)^{1/2}\mathbb{T}_{T-t}^{-1}\boldsymbol{\Gamma}_t$ is uniformly bounded in $t \in [0,T)$. The bound for $(\boldsymbol{\Gamma}_t)_{0\leq t<T}$ easily follows. The bound for $(\gamma_t)_{0\leq t<T}$ is obtained in a similar way since

$$\begin{aligned}\gamma_t &= -A_tB^*[\mathbf{R}(T,t)]^*\int_0^t \boldsymbol{\rho}_s^{-1}B\Sigma_s dW_s \\ &= -A_tB^*\mathbb{T}_{T-t}^{-1}[\hat{\mathbf{R}}^{t,T}(1,0)]^*\mathbb{T}_{T-t}\int_0^t \boldsymbol{\rho}_s^{-1}B\Sigma_s dW_s \\ &= -(T-t)^{-1}A_tB^*[\hat{\mathbf{R}}^{t,T}(1,0)]^*\mathbb{T}_{T-t}\int_0^t \boldsymbol{\rho}_s^{-1}B\Sigma_s dW_s.\end{aligned}$$

The covariance matrix of γ_t (which is in $\mathcal{M}_d(\mathbb{R})$) has the form

$$(T-t)^{-2}A_tB^*[\hat{\mathbf{R}}^{t,T}(1,0)]^*\mathbb{T}_{T-t}\left[\int_0^t \boldsymbol{\rho}_s^{-1}BA_sB^*[\boldsymbol{\rho}_s^{-1}]^*ds\right]\mathbb{T}_{T-t}\hat{\mathbf{R}}^{t,T}(1,0)BA_t^*.$$

By (3.19), we deduce that the covariance matrix is bounded by $C_{1\vee T}(T-t)^{-1}I_d$, I_d standing for the identity matrix of size d. $\square$

4 Proof of the Lower Bound

Here is the core of the proof of the lower bound.

4.1 Nonlinear Controllability

As announced in Point (*ii*) of Subsection 2.2, we first investigate the controllability problem (2.8)-(2.9):

$$\begin{aligned} &I(T,\mathbf{x}_0,\mathbf{y}_0) = \inf\Big\{\int_0^T |\varphi_t|^2 dt : \boldsymbol{\phi}_0 = \mathbf{x}_0, \quad \boldsymbol{\phi}_T = \mathbf{y}_0\Big\}, \\ &\text{with} \quad \dot{\boldsymbol{\phi}}_t = \mathbf{F}(t,\boldsymbol{\phi}_t) + B\varphi_t, \quad 0 \le t \le T; \quad \boldsymbol{\phi}_0 = \mathbf{x}_0, \end{aligned}$$

T and $\mathbf{y}_0$ being fixed as in Subsection 2.1.

As explained in Subsection 2.2, $I(T,\mathbf{x}_0,\mathbf{y}_0)$ is expected to be the typical "off-diagonal" decay of $p(T,\mathbf{x}_0,\mathbf{y}_0)$. In this subsection, we prove that $I(T,\mathbf{x}_0,\mathbf{y}_0)$ is of the same order as $T|\mathbb{T}_T^{-1}(\boldsymbol{\theta}_T(\mathbf{x}_0) - \mathbf{y}_0)|^2$, i.e. of the same order as the "off-diagonal" term in the two-sided bounds in Theorem 1.1. Right below, we first prove a lower bound for $(I(t,\cdot,\cdot))_{0<t\le T}$:

Proposition 4.1 *There exists a constant $C_{4.1} > 0$, depending on* ($\mathbf{A}$) *and T only, such that, for any $0 < t \le T$, for all $\mathbf{x},\mathbf{y} \in \mathbb{R}^{nd}$, $I(t,\mathbf{x},\mathbf{y}) \ge C_{4.1} t |\mathbb{T}_t^{-1}[\boldsymbol{\theta}_t(\mathbf{x}) - \mathbf{y}]|^2$.*

Proof. Fix $t \in (0,T]$ and $\mathbf{x} \in \mathbb{R}^{nd}$ and consider a control $\varphi \in L^2([0,T],\mathbb{R}^d)$ together with the associated trajectory $\boldsymbol{\phi}$ (see (2.8)) under the initial condition $\boldsymbol{\phi}_0 = \mathbf{x}$. Then, we can write

$$\dot{\boldsymbol{\phi}}_s - \dot{\boldsymbol{\theta}}_s(\mathbf{x}) = \mathbf{L}_s\Big(\boldsymbol{\phi}_s - \boldsymbol{\theta}_s(\mathbf{x})\Big) + B\varphi_s, \quad 0 \le s \le t,$$

where $\mathbf{L}_s = \int_0^1 \mathbf{D_x F}(s, \boldsymbol{\theta}_s(\mathbf{x}) + \lambda(\boldsymbol{\phi}_s - \boldsymbol{\theta}_s(\mathbf{x})))d\lambda \in \mathcal{M}_{nd}(\mathbb{R})$ and $\mathbf{D_x F}$ denotes the space derivative of $\mathbf{F}$. By Assumption ($\mathbf{A}$), $|\mathbf{D_x F}(s,\mathbf{z})| \le \kappa$ and, for $2 \le i \le n$, $[\mathbf{D_x F}(s,\mathbf{z})]_{i,i-1} = D_{x_{i-1}} F_i(s, \mathbf{z}^{i-1,n}) \in \mathcal{E}^{i-1}$. We thus interpret the above equation as a controlled linear equation with $\mathbf{0}$ as initial condition and $\boldsymbol{\phi}_t - \boldsymbol{\theta}_t(\mathbf{x})$ as terminal point. By the combination of Propositions 3.1 and 3.4 with the above $(\mathbf{L}_s)_{0\le s\le t}$ and $(\Sigma_s = I_d)_{0\le s\le t}$ (I_d is the identity matrix of size d), we know that there exists a constant $C > 0$, only depending on the parameters quoted in the statement, such that $\int_0^t |\varphi_s|^2 ds \ge Ct|\mathbb{T}_t^{-1}[\boldsymbol{\phi}_t - \boldsymbol{\theta}_t(\mathbf{x})]|^2$. $\square$

The converse bound is more challenging:

Proposition 4.2 *There exists a constant $C_{4.2} \ge 0$, depending on* ($\mathbf{A}$) *and T only, such that, for any $0 < t \le T$ and $\mathbf{x},\mathbf{y} \in \mathbb{R}^{nd}$, we can find a control*

$(\varphi_s)_{0\le s\le t}$ *with values in* $\mathbb{R}^d$ *satisfying*

(1) $\sup_{0\le s\le t}|\varphi_s|^2 \le C_{4.2}|\mathbb{T}_t^{-1}[\boldsymbol{\theta}_t(\mathbf{x})-\mathbf{y}]|^2$
(2) the solution $(\boldsymbol{\phi}_s)_{0\le s\le t}$ *to* (2.8) *associated with* $(\varphi_s)_{0\le s\le t}$ *and with the initial condition* $\boldsymbol{\phi}_0=\mathbf{x}$ *reaches* $\mathbf{y}$ *at time* t, *i.e.* $\boldsymbol{\phi}_t=\mathbf{y}$.

In particular, $I(t,\mathbf{x},\mathbf{y}) \le C_{4.2}t|\mathbb{T}_t^{-1}[\boldsymbol{\theta}_t(\mathbf{x})-\mathbf{y}]|^2$.

Proof. We fix $t\in(0,T]$ and $\mathbf{x},\mathbf{y}\in\mathbb{R}^{nd}$. We are then seeking for a path $(\boldsymbol{\phi}_s)_{0\le s\le t}$, driven by a control $(\varphi_s)_{0\le s\le t}$ of supremum norm less than $t|\mathbb{T}_t^{-1}[\boldsymbol{\phi}_t-\boldsymbol{\theta}_t(\mathbf{x})]|^2$ (up to a multiplicative constant), such that $\boldsymbol{\phi}_0=\mathbf{x}$, $\boldsymbol{\phi}_t=\mathbf{y}$ and $\dot{\boldsymbol{\phi}}_s=\mathbf{F}(s,\boldsymbol{\phi}_s)+B\varphi_s$, $0\le s\le t$. By subtracting the path $(\boldsymbol{\theta}_s(\mathbf{x}))_{0\le s\le t}$ given by (1.4), it is equivalent to find a path from $\mathbf{0}$ to $\mathbf{y}-\boldsymbol{\theta}_t(\mathbf{x})$ for the nonlinear controlled problem driven by the function $\mathbf{F}(s,\boldsymbol{\theta}_s(\mathbf{x})+\cdot)-\mathbf{F}(s,\boldsymbol{\theta}_s(\mathbf{x}))$. In other words, we can assume w.l.o.g. that $\mathbf{x}=\mathbf{0}$ is the initial point, $\mathbf{y}-\boldsymbol{\theta}_t(\mathbf{x})$ is the terminal point and $\mathbf{F}(s,\mathbf{0})=\mathbf{0}$, $0\le s\le t$. We then write

$$\dot{\boldsymbol{\phi}}_s=\mathbf{L}(s,\boldsymbol{\phi}_s)\boldsymbol{\phi}_s+B\varphi_s,\quad 0\le s\le t\ ;\quad \mathbf{L}(s,\boldsymbol{\phi}_s)=\int_0^1\mathbf{D}_{\mathbf{x}}\mathbf{F}(s,\lambda\boldsymbol{\phi}_s)d\lambda. \tag{4.1}$$

The usual way to solve such a controllability problem relies on Schauder's fixed point theorem. The idea is to associate with each $\mathbb{R}^{nd}$-valued path $(\mathbf{z}_s)_{0\le s\le t}$ the linear control problem of type (3.1) driven by $(\mathbf{L}_s^{\mathbf{z}}=\int_0^1\mathbf{D}_{\mathbf{x}}\mathbf{F}(s,\lambda\mathbf{z}_s)d\lambda)_{0\le s\le t}$ and $(\Sigma_s=I_d)_{0\le s\le t}$ (identity matrix of size d) and to seek for a fixed point. We let the reader check that, for each $\mathbf{z}$, the pair $((\mathbf{L}_s^{\mathbf{z}})_{0\le s\le t},I_d)$ satisfies the assumption of Proposition 3.4. In particular, the spectrum of $\mathbf{K}_t^{\mathbf{z}}$, covariance matrix at time t (or equivalently Gram matrix at time t, see Proposition 3.1) associated with the pair $((\mathbf{L}_s^{\mathbf{z}})_{0\le s\le t},I_d)$, is in some $[\gamma^{-1},\gamma]$, γ being positive and independent of $\mathbf{z}$. By Coron [Cor07, Theorem 3.40], the nonlinear controllability problem (4.1) with $\mathbf{0}$ and $\mathbf{y}-\boldsymbol{\theta}_t(\mathbf{x})$ as boundary conditions is then solvable. The exhibited control $(\varphi_s)_{0\le s\le t}$ in the proof is

$$\varphi_s=B^*\Big[\mathbf{R}^{\boldsymbol{\phi}}(t,s)\Big]^*\Big[\mathbf{K}_t^{\boldsymbol{\phi}}\Big]^{-1}\Big(\mathbf{y}-\boldsymbol{\theta}_t(\mathbf{x})\Big),\quad 0\le s\le t,$$

where $\mathbf{R}^{\boldsymbol{\phi}}$ stands for the resolvent associated with the linear equation driven by the coefficient $(\mathbf{L}(s,\boldsymbol{\phi}_s))_{0\le s\le t}$. (Again, as in Proposition 3.1.) This proves (2).

To prove (1) with respect to $C_{4.2}$ depending on T and quantities in (**A**) only, we apply the scaling Lemma 3.6 to $(\mathbf{L}(s,\boldsymbol{\phi}_s),\Sigma_s)_{0\le s\le t}$. (Actually, we should say to a suitable extension $(\mathbf{L}(s,\boldsymbol{\phi}_s),\Sigma_s)_{0\le s\le T}$ of $(\mathbf{L}(s,\boldsymbol{\phi}_s),\Sigma_s)_{0\le s\le t}$ to be in the required framework, say for example $(\mathbf{L}(s\wedge t,\boldsymbol{\phi}_{s\wedge t}),\Sigma_{s\wedge t})_{0\le s\le T}$). Choosing $s_1=1$ and $s_0=s/t$ in the statement of Lemma 3.6, we write $\mathbf{R}^{\boldsymbol{\phi}}(t,s)=\mathbb{T}_t\hat{\mathbf{R}}^{\boldsymbol{\phi},t}(1,s/t)\mathbb{T}_t^{-1}$ and $\mathbf{K}_t^{\boldsymbol{\phi}}=t^{-1}\mathbb{T}_t\hat{\mathbf{K}}_1^{\boldsymbol{\phi},t}\mathbb{T}_t$, so that

$$\varphi_s=B^*t\mathbb{T}_t^{-1}[\hat{\mathbf{R}}^{\boldsymbol{\phi},t}(1,\frac{s}{t})]^*[\hat{\mathbf{K}}_1^{\boldsymbol{\phi},t}]^{-1}\mathbb{T}_t^{-1}(\mathbf{y}-\boldsymbol{\theta}_t(\mathbf{x})).$$

Since the rescaled coefficients satisfy the "right" assumption (3.10) on [0,1], the matrices $\hat{\mathbf{R}}^{\phi}(1, s/t)$ and $[\hat{\mathbf{K}}_1^{\phi}]^{-1}$ are bounded by a constant $C_{4.2}$ depending on T and $(\mathbf{A})$ only. Since $B^* t\mathbb{T}_t^{-1} = B^*$, this completes the proof. $\square$

4.2 Linearization of the Stochastic Control Problem

We now tackle the last part (iii) of the plan detailed in Subsection 2.2. Recall the problem: to get a lower bound for the density $p(T, \mathbf{x}_0, \mathbf{y}_0)$, we think of proving an upper bound for $J_\varepsilon(0, \mathbf{x}_0)$, uniformly in $\varepsilon > 0$. To prove an upper bound for $J_\varepsilon(0, \mathbf{x}_0)$, we think of plugging a specific control $(v_t)_{0 \le t \le T-\varepsilon}$ in the stochastic control representation (2.6). The idea, as written in Subsection 2.2, is to seek $(v_t)_{0 \le t \le T-\varepsilon}$ of the form $v_t = v_t^0 + \varphi_t$, $0 \le t \le T - \varepsilon$, $(\varphi_t)_{0 \le t \le T}$ being given by Proposition 4.2 with $\mathbf{x} = \mathbf{x}_0$ and $\mathbf{y} = \mathbf{y}_0$.

Recall indeed that the controlled SDE associated with $(v_t)_{0 \le t \le T}$ has the form (2.5)

$$d\boldsymbol{\chi}_t = [\mathbf{F}(t, \boldsymbol{\chi}_t) + Bv_t]dt + B\sigma(t, \boldsymbol{\chi}_t)dW_t, \quad 0 \le t \le T - \varepsilon,$$

with the initial condition $\boldsymbol{\chi}_0 = \mathbf{x}_0$. (From now on, both $\mathbf{x}_0$ and ε are fixed once for all). Then, the difference $(\boldsymbol{\chi}_t - \boldsymbol{\phi}_t)_{0 \le t \le T-\varepsilon}$, with $\boldsymbol{\phi}$ as in Proposition 4.2 (i.e. associated with $(\varphi_t)_{0 \le t \le T}$ and $(\mathbf{x}, \mathbf{y}) = (\mathbf{x}_0, \mathbf{y}_0)$), satisfies

$$\begin{aligned} d\big[\boldsymbol{\chi}_t - \boldsymbol{\phi}_t\big] &= [\mathbf{F}(t, \boldsymbol{\chi}_t) - \mathbf{F}(t, \boldsymbol{\phi}_t) + B(v_t - \varphi_t)]dt + B\sigma(t, \boldsymbol{\chi}_t)dW_t \\ &= [\mathbf{F}(t, \boldsymbol{\chi}_t) - \mathbf{F}(t, \boldsymbol{\phi}_t) + Bv_t^0]dt + B\sigma(t, \boldsymbol{\chi}_t)dW_t, \end{aligned}$$

with the initial condition $\boldsymbol{\chi}_0 - \boldsymbol{\phi}_0 = \mathbf{0}$. We are thus reduced to seek for a control $(v_t^0)_{0 \le t \le T-\varepsilon}$, with a reasonable cost, such that $\boldsymbol{\chi}_{T-\varepsilon} - \boldsymbol{\phi}_{T-\varepsilon}$ be close to $\mathbf{y}_0 - \boldsymbol{\phi}_T = \mathbf{0}$ when ε is small. Observing the above equation, we then set, for $0 \le t \le T - \varepsilon$ and $\mathbf{x} \in \mathbb{R}^{nd}$

$$\boldsymbol{\Theta}_t = \boldsymbol{\chi}_t - \boldsymbol{\phi}_t, \quad \boldsymbol{\mathcal{F}}(t, \mathbf{x}) = \mathbf{F}(t, \boldsymbol{\phi}_t + \mathbf{x}) - \mathbf{F}(t, \boldsymbol{\phi}_t), \quad \varsigma(t, \mathbf{x}) = \sigma(t, \boldsymbol{\phi}_t + \mathbf{x}), \tag{4.2}$$

so that the pair $(v_t^0, \boldsymbol{\Theta}_t)_{0 \le t \le T-\varepsilon}$ satisfies

$$d\boldsymbol{\Theta}_t = [\boldsymbol{\mathcal{F}}(t, \boldsymbol{\Theta}_t) + Bv_t^0]dt + B\varsigma(t, \boldsymbol{\Theta}_t)dW_t, \quad \boldsymbol{\Theta}_0 = \mathbf{0}. \tag{4.3}$$

Once again, $(v_t^0)_{0 \le t \le T-\varepsilon}$ is sought to make $\boldsymbol{\Theta}_{T-\varepsilon}$ close to $\mathbf{0}$. In fact, it is reasonable to seek $(v_t^0)_{0 \le t \le T-\varepsilon}$ to make the whole path $(\boldsymbol{\Theta}_t)_{0 \le t \le T-\varepsilon}$ be in the neighborhood of $\mathbf{0}$. Since $\boldsymbol{\mathcal{F}}(t, \mathbf{0}) = \mathbf{0}$, we thus think of seeking $(v_t^0)_{0 \le t \le T-\varepsilon}$ by approximating the above controlled SDE by a linear version driven by the mappings $(\mathbf{L}_t : \mathbf{x} \in \mathbb{R}^{nd} \mapsto \mathbf{D}_{\mathbf{x}}\boldsymbol{\mathcal{F}}(t, \mathbf{0})\mathbf{x})_{0 \le t \le T-\varepsilon}$ and by the diffusion matrix $(\Sigma_t = \varsigma(t, \mathbf{0}))_{0 \le t \le T-\varepsilon}$. (Here, $\mathbf{D}_{\mathbf{x}}\boldsymbol{\mathcal{F}}(t, \mathbf{0})$ stands for the gradient of $\boldsymbol{\mathcal{F}}$.) Indeed, in the linear case, the optimal control is explicitly known: it is given by Proposition 3.7. (See (3.13).) This is the strategy announced in Subsection 2.2.

We here show that it is enough to consider a reduced gradient for the linearization. For $t \in [0, T-\varepsilon]$, the reduced gradient of $\boldsymbol{\mathcal{F}}(t,\cdot)$ at $\mathbf{0}$ is an $nd \times nd$ matrix, denoted by $D\boldsymbol{\mathcal{F}}(t,\mathbf{0})$, with $n-1$ non-zero $d \times d$ blocks only: $[D\boldsymbol{\mathcal{F}}(t,\mathbf{0})]_{i-1,i} = D_{x_{i-1}}\mathcal{F}_i(t,\mathbf{0})$ for $2 \le i \le n$; the other blocks are zero. In other words, only the subdiagonal of $D\boldsymbol{\mathcal{F}}(t,\mathbf{0})$ is non-zero: this gives the same picture as in footnote 6, but with U equal to zero. We emphasize that the blocks on the subdiagonal cannot be zero: since $D_{x_{i-1}}\mathcal{F}_i(t,\mathbf{0}) = D_{x_{i-1}}F_i(t,\boldsymbol{\phi}_t) \in \mathcal{E}_{i-1}$, they are non-degenerate.

By Proposition 3.7, the optimal controlled SDE associated with the coefficients $(\mathbf{L}_t = D\boldsymbol{\mathcal{F}}(t,\mathbf{0}))_{0\le t\le T}$ and $(\Sigma_t = \varsigma(t,\mathbf{0}))_{0\le t\le T}$ may be written

$$\begin{aligned} d\boldsymbol{\Gamma}_t &= D\boldsymbol{\mathcal{F}}(t,\mathbf{0})\boldsymbol{\Gamma}_t dt + B\gamma_t dt + B\varsigma(t,\mathbf{0})dW_t, \quad \boldsymbol{\Gamma}_0 = \mathbf{0}, \\ &\text{where } \gamma_t = -A_t B^*[\mathbf{R}(T,t)]^* \int_0^t \boldsymbol{\rho}_s^{-1} B\varsigma(s,\mathbf{0})dW_s, \end{aligned} \tag{4.4}$$

$A_t = \varsigma\varsigma^*(t,\mathbf{0}) = a(t,\boldsymbol{\phi}_t)$, $(\mathbf{R}(t,s))_{0\le t,s\le T}$ is the resolvent associated with $(\mathbf{L}_t = D\boldsymbol{\mathcal{F}}(t,\mathbf{0}))_{0\le t\le T}$ and $\boldsymbol{\rho}_t = [\int_t^T \mathbf{R}(t,s)BA_sB^*[\mathbf{R}(t,s)]^* ds][\mathbf{R}(T,t)]^*$, $0 \le t \le T$.

On this model, we choose

Proposition 4.3 *Under the notation of Proposition 3.7, define* $\boldsymbol{\Theta}$ *in* (4.3) *as the solution of the SDE with*

$$v_t^0 = -A_t B^*[\mathbf{R}(T,t)]^* \int_0^t \boldsymbol{\rho}_s^{-1}\Big[[\boldsymbol{\mathcal{F}}(s,\boldsymbol{\Theta}_s) - D\boldsymbol{\mathcal{F}}(s,\mathbf{0})\boldsymbol{\Theta}_s]ds + B\varsigma(s,\boldsymbol{\Theta}_s)dW_s\Big],$$

$0 \le t \le T-\varepsilon$. *Fix* $\mu > 0$. *Then, there exists two constants* $c_{4.3}(\mu) > 0$ *and* $C_{4.3}(\mu) > 0$, *only depending on* ($\mathbf{A}$), *such that for* $T \le c_{4.3}(\mu)$,

$$\mathbb{P}\left\{\forall t \in [0,T-\varepsilon], \begin{array}{l} (T-t)|\mathbb{T}_{T-t}^{-1}(\boldsymbol{\Theta}_t - \boldsymbol{\Gamma}_t)| \le C_{4.3}(\mu)(T-t)^{\frac{1}{2}+\frac{\eta}{8}} \\ |v_t^0 - \gamma_t| \le C_{4.3}(\mu)(T-t)^{-\frac{1}{2}+\frac{\eta}{8}} \end{array}\right\} \ge 1-\mu. \tag{4.5}$$

We emphasize that the above estimate is crucial: it shows that $\boldsymbol{\Theta}_{T-\varepsilon} - \boldsymbol{\Gamma}_{T-\varepsilon}$ is close to $\mathbf{0}$ with a large probability when ε is small. Since $\boldsymbol{\Gamma}_{T-\varepsilon}$ itself is close to $\mathbf{0}$ (see Lemma 3.9), this shows that $\boldsymbol{\Theta}_{T-\varepsilon}$ is also close to $\mathbf{0}$ with a large probability when ε is small.

Proof. In the whole proof, we assume $T \le 1$.

First Step. We first note that the SDE obtained by plugging the definition of $(v_t^0)_{0\le t\le T-\varepsilon}$ into (4.3) is well-posed. (Note that v^0 depends on $\boldsymbol{\Theta}$ itself.) The well-posedness follows from (3.14) and from the Lipschitz property of $\boldsymbol{\mathcal{F}}$ (or equivalently of $\mathbf{F}$): we refer to [RY99, Thm. 2.1, Chap. IX] for the unique solvability of such a non-Markovian SDE. By (3.17), we also know that $\dot{\boldsymbol{\rho}}_t = \mathbf{L}_t \boldsymbol{\rho}_t - BA_t B^* [\mathbf{R}(T,t)]^*$, $0 \le t < T$. By variation of parameters, we thus obtain the equivalent of (3.15) for $\boldsymbol{\Theta}$: for $0 \le t \le T - \varepsilon$,

$$\boldsymbol{\Theta}_t = \boldsymbol{\rho}_t \int_0^t \boldsymbol{\rho}_s^{-1} \Big[[\boldsymbol{\mathcal{F}}(s, \boldsymbol{\Theta}_s) - D\boldsymbol{\mathcal{F}}(s, \mathbf{0}) \boldsymbol{\Theta}_s] ds + B\varsigma(s, \boldsymbol{\Theta}_s) dW_s \Big].$$

Substracting (3.15) with $\Sigma_s = \varsigma(s, 0)$, we obtain, for $0 \le t \le T - \varepsilon$,

$$\begin{aligned} &\boldsymbol{\Theta}_t - \boldsymbol{\Gamma}_t \\ &= \boldsymbol{\rho}_t \int_0^t \boldsymbol{\rho}_s^{-1} \Big[[\boldsymbol{\mathcal{F}}(s, \boldsymbol{\Theta}_s) - D\boldsymbol{\mathcal{F}}(s, \mathbf{0}) \boldsymbol{\Theta}_s] ds + B[\varsigma(s, \boldsymbol{\Theta}_s) - \varsigma(s, \mathbf{0})] dW_s \Big] \\ &= \boldsymbol{\rho}_t \int_0^t \boldsymbol{\rho}_s^{-1} [\Delta \boldsymbol{\mathcal{F}}_s ds + B \Delta \varsigma_s dW_s], \end{aligned} \tag{4.6}$$

with

$$\Delta \boldsymbol{\mathcal{F}}_s = \boldsymbol{\mathcal{F}}(s, \boldsymbol{\Theta}_s) - D\boldsymbol{\mathcal{F}}(s, \mathbf{0}) \boldsymbol{\Theta}_s, \quad \Delta \varsigma_s = \varsigma(s, \boldsymbol{\Theta}_s) - \varsigma(s, \mathbf{0}).$$

We also set for all $t \in [0, T - \varepsilon]$,

$$\begin{aligned} &\hat{\boldsymbol{\Gamma}}_t = (T-t) \mathbb{T}_{T-t}^{-1} \boldsymbol{\Gamma}_t, \\ &\hat{\boldsymbol{\Theta}}_t = (T-t) \mathbb{T}_{T-t}^{-1} \boldsymbol{\Theta}_t, \\ &\mathbf{E}_t = \hat{\boldsymbol{\Theta}}_t - \hat{\boldsymbol{\Gamma}}_t = (T-t) \mathbb{T}_{T-t}^{-1} (\boldsymbol{\Theta}_t - \boldsymbol{\Gamma}_t), \\ &\Delta \hat{\boldsymbol{\mathcal{F}}}_t = (T-t)^2 \mathbb{T}_{T-t}^{-1} \Delta \boldsymbol{\mathcal{F}}_t. \end{aligned} \tag{4.7}$$

By (3.18), we know that $\boldsymbol{\rho}_t = (T-t)^{-1} \mathbb{T}_{T-t} \hat{\mathbf{R}}^{t,T}(0,1) \hat{\mathbf{K}}_1^{t,T} \mathbb{T}_{T-t}$, where $\hat{\mathbf{R}}^{t,T}(0,1)$ and $\hat{\mathbf{K}}_1^{t,T}$ stand for suitable rescaled matrices for which (upper and lower) bounds, only depending on $(\mathbf{A})$ (and not on T), are known, uniformly in $0 \le t \le T$ (see Lemma 3.6). Since $T \le 1$, we indeed emphasize that the rescaled coefficients in the definition of $\hat{\mathbf{R}}^{t,T}(0,1)$ and $\hat{\mathbf{K}}_1^{t,T}$ satisfy the same assumptions as the initial non-rescaled coefficients (see (3.10)); this explains why the bounds for $\hat{\mathbf{R}}^{t,T}(0,1)$ and $\hat{\mathbf{K}}_1^{t,T}$ are independent of T. By (4.6) and (4.7),

$$\mathbf{E}_t = \hat{\mathbf{R}}^{t,T}(0,1) \hat{\mathbf{K}}_1^{t,T} \mathbb{T}_{T-t} \int_0^t \boldsymbol{\rho}_s^{-1} [\Delta \boldsymbol{\mathcal{F}}_s ds + B \Delta \varsigma_s dW_s], \tag{4.8}$$

so that

$$\begin{aligned} |\mathbf{E}_t| \le C \bigg| \int_0^t & \mathbb{T}_{T-t} \mathbb{T}_{T-s}^{-1} [\hat{\mathbf{R}}^{s,T}(0,1) \hat{\mathbf{K}}_1^{s,T}]^{-1} \\ & \times (T-s) \mathbb{T}_{T-s}^{-1} [\Delta \boldsymbol{\mathcal{F}}_s ds + B \Delta \varsigma_s dW_s] \bigg|, \end{aligned} \tag{4.9}$$

where C is a constant depending on $(\mathbf{A})$ only (whose value may vary from line to line). In particular, C is independent of T. Since $(T-s) \mathbb{T}_{T-s}^{-1} B = B$, we write $(T-s) \mathbb{T}_{T-s}^{-1} [\Delta \boldsymbol{\mathcal{F}}_s ds + B \Delta \varsigma_s dW_s] = [(T-s)^{-1} \Delta \hat{\boldsymbol{\mathcal{F}}}_s ds + B \Delta \varsigma_s dW_s]$. (See (4.7) for the definition of $\Delta\hat{\boldsymbol{\mathcal{F}}}$.) Since $|\mathbb{T}_{T-t}\mathbb{T}_{T-s}^{-1}\mathbf{x}|^2 = \sum_{i=1}^n (T-t)^{2i}(T-s)^{-2i}|x_i|^2$ for $\mathbf{x} = (x_1, \dots, x_n) \in (\mathbb{R}^d)^n$, we obtain

$$|\mathbf{E}_t| \le C\sum_{i=1}^n (T-t)^i \left[\left|\int_0^t \frac{N_s^i \Delta\varsigma_s}{(T-s)^i} dW_s\right| + \int_0^t \frac{|\Delta\hat{\boldsymbol{\mathcal{F}}}_s|}{(T-s)^{i+1}} ds\right], \tag{4.10}$$

where N_s^i stands for the $d \times d$ block of index $(i,1)$ of the $nd \times nd$ matrix $[\hat{\mathbf{R}}^{s,T}(0,1)\hat{\mathbf{K}}_1^{s,T}]^{-1}$, i.e. $N_s^i = [[\hat{\mathbf{R}}^{s,T}(0,1)\hat{\mathbf{K}}_1^{s,T}]^{-1}]_{i,1}$.

Second Step. To prove the bound on $\mathbf{E}$, we introduce the stopping time

$$\tau = \inf\{t \in [0, T-\varepsilon] : |\mathbf{E}_t| \ge (T-t)^{\frac{1}{2}+\frac{\eta}{8}}\} \quad (\inf\emptyset = T-\varepsilon). \tag{4.11}$$

Saying that $|\mathbf{E}_t|$ is less than $(T-t)^{\frac{1}{2}+\frac{\eta}{8}}$ with large probability is the same as saying that $\tau = T-\varepsilon$ with large probability. To get rid of the martingale terms in (4.10), we introduce the set

$$\mathcal{A} = \bigcap_{i=1}^n \Big\{\forall t \in [0,\tau] : \ (T-t)^i \Big|\int_0^t (T-s)^{-i} N_s^i \Delta\varsigma_s dW_s\Big| \le \alpha(T-t)^{\frac{1}{2}+\frac{\eta}{4}}\Big\}, \tag{4.12}$$

the constant α being chosen later to make $\mathbb{P}(\mathcal{A})$ be as large as possible. On $\mathcal{A}$, from (4.10)

$$|\mathbf{E}_t| \le C\alpha(T-t)^{\frac{1}{2}+\frac{\eta}{4}} + C\sum_{i=1}^n (T-t)^i \int_0^t \frac{|\Delta\hat{\boldsymbol{\mathcal{F}}}_s|}{(T-s)^{i+1}} ds. \tag{4.13}$$

We now use the regularity of $\boldsymbol{\mathcal{F}}$ (or equivalently of $\mathbf{F}$) to bound $|\Delta\hat{\boldsymbol{\mathcal{F}}}_s|$. For $j \ge 2$, we write the j^{th} block of $\Delta\boldsymbol{\mathcal{F}}_s$ as

$$\begin{aligned}(\Delta\boldsymbol{\mathcal{F}}_s)_j &= [\mathcal{F}_j(s, \boldsymbol{\Theta}_s^{j-1,n}) - \mathcal{F}_j(s, \Theta_s^{j-1}, 0, \dots, 0)] \\ &\quad + [\mathcal{F}_j(s, \Theta_s^{j-1}, 0, \dots, 0) - \mathcal{F}_j(s, \mathbf{0}^{j-1,n}) - D_{x_{j-1}}\boldsymbol{\mathcal{F}}_j(s, \mathbf{0}^{j-1,n})\Theta_s^{j-1}],\end{aligned}$$

since $\mathcal{F}_j(s, \mathbf{0}^{j-1,n}) = 0$. Here, $\mathcal{F}_j$ is the j^{th} coordinate of $\boldsymbol{\mathcal{F}}$. Using the Lipschitz property of $\mathcal{F}_j$ for the first part and expanding the second part with Taylor's integral formula at order 1 recalling also that, under (**A**), $D_{x_{j-1}}\mathcal{F}_j$ is η-Hölder continuous in the $(j-1)^{\text{th}}$ space variable, $\eta \in (0,1]$, and $\mathcal{F}_1(s, \mathbf{0}) = 0$, we obtain (see (4.7) for the definition of $\Delta\hat{\boldsymbol{\mathcal{F}}}$ in terms of $\Delta\boldsymbol{\mathcal{F}}$):

$$\begin{aligned}|\Delta\hat{\boldsymbol{\mathcal{F}}}_s| &\le C\Big((T-s)[\mathcal{F}_1(s, \boldsymbol{\Theta}_s) - \mathcal{F}_1(s, \mathbf{0})] \\ &\qquad + \sum_{j=2}^n (T-s)^{-(j-2)}\Big[|\mathcal{F}_j(s, \boldsymbol{\Theta}_s^{j-1,n}) - \mathcal{F}_j(s, \Theta_s^{j-1}, 0, \dots, 0)| \\ &\qquad + |\mathcal{F}_j(s, \Theta_s^{j-1}, 0, \dots, 0) - \mathcal{F}_j(s, \mathbf{0}^{j-1,n}) - D_{x_{j-1}}\mathcal{F}_j(s, \mathbf{0}^{j-1,n})\Theta_s^{j-1}|\Big]\Big) \\ &\le C\sum_{j=1}^n (T-s)^{-(j-2)}\Big[|\Theta_s^{j-1}|^{1+\eta} + |\boldsymbol{\Theta}_s^{j,n}|\Big] \\ &\le C\Big(|(T-s)\mathbb{T}_{T-s}^{-1}\boldsymbol{\Theta}_s|^{1+\eta} + (T-s)^2|\mathbb{T}_{T-s}^{-1}\boldsymbol{\Theta}_s|\Big),\end{aligned}$$

with the convention $\Theta^0 = 0$.

Having in mind the notations $\hat{\mathbf{\Theta}}_s = (T-s)\mathbb{T}_{T-s}^{-1}\mathbf{\Theta}_s$ and $\hat{\mathbf{\Gamma}}_s = (T-s)\mathbb{T}_{T-s}^{-1}\mathbf{\Gamma}_s$ and the decomposition $\hat{\mathbf{\Theta}}_s = \mathbf{E}_s + \hat{\mathbf{\Gamma}}_s$, we obtain

$$\begin{aligned}|\Delta\hat{\boldsymbol{\mathcal{F}}}_s| &\le C\Big[|(T-s)\mathbb{T}_{T-s}^{-1}\mathbf{\Theta}_s|^{1+\eta} + |(T-s)^2\mathbb{T}_{T-s}^{-1}\mathbf{\Theta}_s|\Big]\\ &\le C\Big[|\mathbf{E}_s + \hat{\mathbf{\Gamma}}_s|^{1+\eta} + (T-s)|\mathbf{E}_s + \hat{\mathbf{\Gamma}}_s|\Big]\\ &\le C\Big[(T-s)^{\frac{\eta+1}{\eta}} + |\mathbf{E}_s|^{1+\eta} + |\hat{\mathbf{\Gamma}}_s|^{1+\eta}\Big],\end{aligned}$$

using usual Young and convexity inequalities to derive the last bound. Plugging this bound into (4.13), we obtain, on $\mathcal{A}$,

$$\begin{aligned}|\mathbf{E}_t| &\le C\alpha(T-t)^{\frac12+\frac\eta4}\\ &\quad + C\sum_{i=1}^n (T-t)^{\frac12+\frac\eta4}\int_0^t (T-t)^{i-(\frac12+\frac\eta4)}\frac{(T-s)^{\frac{\eta+1}{\eta}} + |\mathbf{E}_s|^{1+\eta} + |\hat{\mathbf{\Gamma}}_s|^{1+\eta}}{(T-s)^{i+1}}ds.\end{aligned}$$

Noting that $(T-t)^{i-(\frac12+\frac\eta4)} \le (T-s)^{i-(\frac12+\frac\eta4)}$ for $i \ge 1$ and $0 \le s \le t \le T$, we have $\int_0^t (T-t)^{i-(\frac12+\frac\eta4)}(T-s)^{\frac{\eta+1}{\eta}}(T-s)^{-(i+1)}ds \le C$ (since $T \le 1$) and

$$|\mathbf{E}_t| \le C(1+\alpha)(T-t)^{\frac12+\frac\eta4} + C(T-t)^{\frac12+\frac\eta4}\int_0^t \frac{|\mathbf{E}_s|^{1+\eta} + |\hat{\mathbf{\Gamma}}_s|^{1+\eta}}{(T-s)^{\frac32+\frac\eta4}}ds.$$

To get rid of the randomness of $\hat{\mathbf{\Gamma}}$, we introduce a new “good event”:

$$\mathcal{B} = \Big\{\forall t \in [0, T-\varepsilon] : \int_0^t (T-s)^{-(\frac32+\frac\eta4)}|\hat{\mathbf{\Gamma}}_s|^{1+\eta}ds \le \beta\Big\}, \tag{4.14}$$

β being chosen later on to make $\mathbb{P}(\mathcal{B})$ be as large as possible. On $\mathcal{A}\cap\mathcal{B}$, we finally have for all $t \in [0, T-\varepsilon]$,

$$|\mathbf{E}_t| \le C(1+\alpha+\beta)(T-t)^{\frac12+\frac\eta4} + C(T-t)^{\frac12+\frac\eta4}\int_0^t \frac{|\mathbf{E}_s|^{1+\eta}}{(T-s)^{\frac32+\frac\eta4}}ds. \tag{4.15}$$

For $t = \tau\wedge(T-\varepsilon)$, we have by (4.11)

$$\int_0^{\tau\wedge(T-\varepsilon)} \frac{|\mathbf{E}_s|^{1+\eta}}{(T-s)^{\frac32+\frac\eta4}}ds \le \int_0^\tau \frac{(T-s)^{\frac12+\frac\eta2}}{(T-s)^{\frac32+\frac\eta4}}ds \le C.$$

Choosing $t = \tau\wedge(T-\varepsilon)$ in (4.15), we deduce (from (4.11) again) $\tau < T-\varepsilon \Rightarrow (T-\tau)^{\frac12+\frac\eta8} \le C(\alpha,\beta)(T-\tau)^{\frac12+\frac\eta4}$. (The constant $C(\alpha,\beta)$ only depends on α, β and (**A**). In particular, it is independent of T.) For T small enough, this is impossible: this proves that

$$\mathcal{A}\cap\mathcal{B} \subset \{\forall t \in [0,T-\varepsilon],\ |\mathbf{E}_t| \le (T-t)^{\frac12+\frac\eta8}\} \quad \text{if} \quad T \le c(\alpha,\beta), \tag{4.16}$$

$c(\alpha,\beta)$ being a positive constant only depending on α, β and $(\mathbf{A})$.

Third Step. To make the strategy relevant, it remains to estimate $\mathbb{P}(\mathcal{A}\cap\mathcal{B})$. We first prove that $\mathbb{P}(\mathcal{B}^{\complement})$ tends to zero as β tends $+\infty$. Indeed, by (4.14) and by Lemma 3.9

$$
\begin{aligned}
\mathbb{P}\big(\mathcal{B}^{\complement}\big) &= \mathbb{P}\Big\{\int_0^{T-\varepsilon}(T-s)^{-(\frac{3}{2}+\frac{\eta}{4})}|\hat{\mathbf{\Gamma}}_s|^{1+\eta}ds > \beta\Big\}\\
&\le \beta^{-1}\mathbb{E}\int_0^{T-\varepsilon}(T-s)^{-(\frac{3}{2}+\frac{\eta}{4})}|\hat{\mathbf{\Gamma}}_s|^{1+\eta}ds\\
&= \beta^{-1}\mathbb{E}\int_0^{T-\varepsilon}(T-s)^{-1+\frac{\eta}{4}}|(T-s)^{\frac{1}{2}}\mathbb{T}_{T-s}^{-1}\mathbf{\Gamma}_s|^{1+\eta}ds\\
&\le C\beta^{-1}\int_0^{T-\varepsilon}(T-s)^{-1+\frac{\eta}{4}}ds \le C\beta^{-1}.
\end{aligned}
\tag{4.17}
$$

(We emphasize that C may be chosen independently of T, as done above, since $T\le 1$.) Therefore, $\mathbb{P}(\mathcal{B}^{\complement})$ can be made as small as desired by choosing β large enough.

It now remains to show that $\lim_{\alpha\to 0}\mathbb{P}(\mathcal{A}^{\complement})=0$. Without loss of generality, we perform the proof for $d=1$ (i.e. when the martingale terms in the definition of $\mathcal{A}$, see (4.12), are one-dimensional). Otherwise, we have to make the proof for the coordinates of each martingale term. The proof relies on Lemma A.2 in Appendix. We apply it to each $((T-t)^i\int_0^{t\wedge\tau}(T-s)^{-i}N_s^i\Delta\varsigma_s dW_s)_{0\le t\le T-\varepsilon}$, $1\le i\le n$. Indeed, it is clear that N_s^i is bounded by some C depending on $(\mathbf{A})$ only and that $\Delta\varsigma_s=\varsigma(s,\mathbf{\Theta}_s)-\varsigma(s,\mathbf{0})$ is bounded. In fact, $|\Delta\varsigma_s|\le C|\mathbf{\Theta}_s|^\eta\le C(|\mathbf{E}_s|^\eta+|\hat{\mathbf{\Gamma}}_s|^\eta)$. Therefore, applying Lemma A.2 for some $\alpha'>0$ (to be chosen in terms of α) and for $\mu=1+\eta/2$, we have with probability greater than $1-\exp(-(\alpha')^2)$, for any $t\in[0,T-\varepsilon]$,

$$
\begin{aligned}
&(T-t)^i\Big|\int_0^{t\wedge\tau}(T-s)^{-i}N_s^i\Delta\varsigma_s dW_s\Big|\\
&\le C(T-t)^{\frac{1}{2}+\frac{\eta}{4}}\Big(\int_0^\tau\frac{|\mathbf{E}_s|^{2\eta}+|\hat{\mathbf{\Gamma}}_s|^{2\eta}}{(T-s)^{1+\frac{\eta}{2}}}ds+\alpha'\Big)^{\frac{1}{2}}\exp\Big(C\alpha'\int_0^\tau\frac{|\mathbf{E}_s|^{2\eta}+|\hat{\mathbf{\Gamma}}_s|^{2\eta}}{(T-s)^{1+\frac{\eta}{2}}}ds\Big).
\end{aligned}
\tag{4.18}
$$

Using Hölder inequality on $[0,T)$ equipped with the (finite) measure $(T-s)^{-1+\frac{\eta}{2}}ds$, it holds on $\mathcal{B}$ (see (4.14))

$$
\begin{aligned}
\int_0^\tau\frac{|\hat{\mathbf{\Gamma}}_s|^{2\eta}}{(T-s)^{1+\frac{\eta}{2}}}ds &= \int_0^\tau(T-s)^{-1+\frac{\eta}{2}}\frac{|\hat{\mathbf{\Gamma}}_s|^{2\eta}}{(T-s)^\eta}ds\\
&\le C\Big(\int_0^\tau(T-s)^{-1+\frac{\eta}{2}}\frac{|\hat{\mathbf{\Gamma}}_s|^{1+\eta}}{(T-s)^{\frac{1}{2}+\frac{\eta}{2}}}ds\Big)^{\frac{2\eta}{1+\eta}}\\
&= C\Big(\int_0^\tau\frac{|\hat{\mathbf{\Gamma}}_s|^{1+\eta}}{(T-s)^{\frac{3}{2}}}ds\Big)^{\frac{2\eta}{1+\eta}}\le C\beta.
\end{aligned}
\tag{4.19}
$$

(As above, C may be chosen independently of T since $T \le 1$.) Moreover, by definition of τ (see (4.11)),

$$\int_0^\tau \frac{|\mathbf{E}_s|^{2\eta}}{(T-s)^{1+\frac{\eta}{2}}} ds \le \int_0^\tau \frac{(T-s)^{\eta+\frac{\eta^2}{4}}}{(T-s)^{1+\frac{\eta}{2}}} ds \le C. \tag{4.20}$$

By (4.19) and (4.20), we deduce that, on $\mathcal{B}$, the right-hand side in (4.18) is less than $C(\alpha', \beta)(T-t)^{\frac{1}{2}+\frac{\eta}{4}}$ for some constant $C(\alpha', \beta)$ depending on α', β and $(\mathbf{A})$ only. (Once again, it is independent of T.) We deduce that, with probability greater than $1 - \exp(-(\alpha')^2) - \mathbb{P}(\mathcal{B}^{\complement})$, for any $t \in [0, T-\varepsilon]$, $(T-t)^i | \int_0^{t\wedge\tau} (T-s)^{-i} N_s^i \Delta\varsigma_s dW_s| \le C(\alpha', \beta)(T-t)^{\frac{1}{2}+\frac{\eta}{4}}$. By (4.12) and (4.17), we obtain that $\mathbb{P}(\mathcal{A}^{\complement}) \le n(\exp(-(\alpha')^2) + C\beta^{-1})$ whenever $\alpha \ge C(\alpha', \beta)$. This proves that $\mathbb{P}(\mathcal{A}^{\complement})$ can be made as small as desired by choosing α large enough.

Fourth Step. We now prove the first line in (4.5). For $\mu > 0$, we can choose α and β large enough such that $\mathbb{P}(\mathcal{A} \cap \mathcal{B}) \ge 1 - \mu$. By (4.7) and (4.16), we deduce that $\mathbb{P}\{\forall t \in [0, T-\varepsilon],\ (T-t)|\mathbb{T}_{T-t}^{-1}(\boldsymbol{\Gamma}_t - \boldsymbol{\Theta}_t)| \le (T-t)^{\frac{1}{2}+\frac{\eta}{8}}\} \ge 1 - \mu$ for $T \le c(\alpha, \beta)$.

It now remains to prove that $|v_t^0 - \gamma_t|$ is controlled by $(T-t)^{-\frac{1}{2}+\frac{\eta}{8}}$ on an event of large probability. By definition of v^0 and γ (see (4.4) together with the statement of Proposition 4.3), we have

$$\begin{aligned} v_t^0 - \gamma_t &= -A_t B^* [\mathbf{R}(T,t)]^* \int_0^t \boldsymbol{\rho}_s^{-1} \Big[\Delta \boldsymbol{\mathcal{F}}_s ds + B \Delta\varsigma_s dW_s \Big] \\ &= -A_t B^* \mathbb{T}_{T-t}^{-1} [\hat{\mathbf{R}}^{t,T}(1,0)]^* \mathbb{T}_{T-t} \int_0^t \boldsymbol{\rho}_s^{-1} [\Delta \boldsymbol{\mathcal{F}}_s ds + B \Delta\varsigma_s dW_s], \end{aligned}$$

since $[\mathbf{R}(T,t)]^* = \mathbb{T}_{T-t}^{-1}[\hat{\mathbf{R}}^{t,T}(1,0)]^* \mathbb{T}_{T-t}$. (See the beginning of the proof of Proposition 3.7 for the scaling argument deriving from Lemma 3.6.) Since $B^* \mathbb{T}_{T-t}^{-1} = (T-t)^{-1} B^*$, we obtain

$$|v_t^0 - \gamma_t| \le C(T-t)^{-1} \Big| \mathbb{T}_{T-t} \int_0^t \boldsymbol{\rho}_s^{-1} [\Delta \boldsymbol{\mathcal{F}}_s ds + B \Delta\varsigma_s dW_s] \Big|.$$

Up to the term $C(T-t)^{-1}$, the above right-hand side is exactly the same as the right-hand side in (4.8)–(4.9). This explains why the exponent in the growth of $|v_t^0 - \gamma_t|$ is equal to the exponent in the growth of $|\mathbf{E}_t|$ minus 1 on the event $\mathcal{A} \cap \mathcal{B}$. The final estimate is easily deduced. $\square$

4.3 Conclusion in Short Time

(All the notations introduced in the beginning of the previous subsection still hold. Moreover, as in the proof of Proposition 4.3, T is assumed to be less than 1.)

We now prove the lower bound for the density when T is small enough. As already explained in Subsection 4.2, the idea is to provide an upper bound for $J_\varepsilon(0,\mathbf{x}_0)$ by plugging $v_t = v_t^0 + \varphi_t$, $0 \le t \le T-\varepsilon$, in (2.7), $(\varphi_t)_{0\le t\le T}$ as in the previous subsection (i.e. being given by Proposition 4.2 with $\mathbf{x} = \mathbf{x}_0$ and $\mathbf{y} = \mathbf{y}_0$) and $(v_t^0)_{0\le t\le T}$ being given by Proposition 4.3. We also specify the choice for η_ε: we choose $\eta_\varepsilon(\mathbf{x}) = q(T-\varepsilon, T, \mathbf{x}-\boldsymbol{\phi}_{T-\varepsilon}, \mathbf{0})$ where $(q(s,t,\mathbf{x},\mathbf{y}))_{0\le s<t\le T;\mathbf{x},\mathbf{y}\in\mathbb{R}^{nd}}$ stands for the transition density of $(\mathbf{G}_t)_{0\le t\le T}$ in Proposition 3.1 with $(\mathbf{L}_t = D\boldsymbol{\mathcal{F}}(t,\mathbf{0}))_{0\le t\le T}$ and $(\Sigma_t = \varsigma(t,\mathbf{0}))_{0\le t\le T}$. (Compare with the controlled equation (4.4).)

We notice that such an η_ε satisfies the assumption required in Subsection 2.1. Clearly, $q(T-\varepsilon, T, \cdot - \boldsymbol{\phi}_{T-\varepsilon}, \mathbf{0})$ weakly converges towards the Dirac mass at $\mathbf{y}_0$ since $\boldsymbol{\phi}_{T-\varepsilon} \to \mathbf{y}_0$ as ε vanishes. Moreover, $q(T-\varepsilon, T, \cdot, \mathbf{0})$ is positive by Proposition 3.3 (applied between $T-\varepsilon$ and T instead of 0 and t). Finally, by Proposition 3.3 again, $q(T-\varepsilon, T, \mathbf{y}-\boldsymbol{\phi}_{T-\varepsilon}, \mathbf{0}) \le C\varepsilon^{-n^2d/2}\exp(-C^{-1}\varepsilon^{-1}|\mathbb{T}_\varepsilon^{-1}\mathbf{R}(T, T-\varepsilon)(\mathbf{y}-\boldsymbol{\phi}_{T-\varepsilon})|^2)$ for ε small enough and for C independent of ε and $\mathbf{y}$: (2.2) follows.

First Step. From (2.7) with $(v_t = v_t^0 + \varphi_t)_{0\le t\le T-\varepsilon}$, we write, for some constant $C>0$, depending on $(\mathbf{A})$ only (as in the previous subsection, C is independent of T),

$$
\begin{aligned}
J_\varepsilon(0,\mathbf{x}_0) \le{}& -\ln[\eta_\varepsilon(\boldsymbol{\chi}_{T-\varepsilon})] + \frac{1}{2}\int_0^{T-\varepsilon} \langle a^{-1}(t,\boldsymbol{\phi}_t)v_t^0, v_t^0\rangle dt + C\int_0^{T-\varepsilon} |\varphi_t|^2 dt \\
&+ C\int_0^{T-\varepsilon} |\varphi_t||v_t^0| dt + C\int_0^{T-\varepsilon} |\boldsymbol{\chi}_t - \boldsymbol{\phi}_t|^\eta |v_t^0|^2 dt \\
&- \frac{1}{2}\int_0^{T-\varepsilon} |\sigma^{-1}(t,\boldsymbol{\chi}_t)[v_t^* - v_t]|^2 dt + \int_0^{T-\varepsilon} \langle \sigma^{-1}(t,\boldsymbol{\chi}_t)v_t^*, dW_t\rangle.
\end{aligned}
$$

By Proposition 4.2, the cost of φ is bounded by $CT|\mathbb{T}_T^{-1}[\boldsymbol{\theta}_T(\mathbf{x}_0) - \mathbf{y}_0]|^2$. (Here again, the constant T may be chosen independently of T since $T \le 1$: Proposition 4.2 says that the constant $C_{4.2}$ is uniform on every bounded interval. In other words, we can choose the constant $C_{4.2}$ associated with the interval $[0,1]$.) By plugging (3.11) (with $(\Sigma_t = \varsigma(t,\mathbf{0}) = \sigma(t,\boldsymbol{\phi})_t)_{0\le t\le T}$ and $(A_t = \varsigma\varsigma^*(t,\mathbf{0}))_{0\le t\le T}$) into the above equation, we deduce

$$
J_\varepsilon(0,\mathbf{x}_0) \le CT|\mathbb{T}_T^{-1}[\boldsymbol{\theta}_T(\mathbf{x}_0) - \mathbf{y}_0]|^2 - \ln(q(T,\mathbf{0},\mathbf{0})) + R_{T-\varepsilon}, \tag{4.21}
$$

where

$$
\begin{aligned}
&R_{T-\varepsilon}\\
&= -\ln[\eta_\varepsilon(\boldsymbol{\chi}_{T-\varepsilon})] + \ln[q(T-\varepsilon, T, \boldsymbol{\Gamma}_{T-\varepsilon}, \mathbf{0})]\\
&\quad + \frac{1}{2}\int_0^{T-\varepsilon} \Big[\langle a^{-1}(t,\boldsymbol{\phi}_t)v_t^0, v_t^0\rangle - \langle a^{-1}(t,\boldsymbol{\phi}_t)\gamma_t,\gamma_t\rangle\Big]dt\\
&\quad + C\int_0^{T-\varepsilon} |\varphi_t||v_t^0|dt\\
&\quad + C\int_0^{T-\varepsilon} |\boldsymbol{\chi}_t - \boldsymbol{\phi}_t|^\eta |v_t^0|^2 dt\\
&\quad - \frac{1}{2}\int_0^{T-\varepsilon} |\sigma^{-1}(t,\boldsymbol{\chi}_t)[v_t^* - v_t]|^2 dt + \int_0^{T-\varepsilon} \langle \sigma^{-1}(t,\boldsymbol{\chi}_t)[v_t^* - v_t], dW_t\rangle\\
&\quad + \int_0^{T-\varepsilon} \langle \sigma^{-1}(t,\boldsymbol{\chi}_t)v_t - \sigma^{-1}(t,\boldsymbol{\phi}_t)\gamma_t, dW_t\rangle\\
&:= R^1_{T-\varepsilon} + R^2_{T-\varepsilon} + R^3_{T-\varepsilon} + R^4_{T-\varepsilon} + R^5_{T-\varepsilon} + R^6_{T-\varepsilon}.
\end{aligned}
\tag{4.22}
$$

Second Step. We first treat the last terms $R^5_{T-\varepsilon}$ and $R^6_{T-\varepsilon}$. By Lemma A.1 (with $\beta = 1$), we can choose $\bar\alpha$ large enough so that the two inequalities below be true on a set $\bar{\mathcal{A}}$ with $\mathbb{P}(\bar{\mathcal{A}}) \geq 1 - 2\exp(-\bar\alpha)$:

$$
R^5_{T-\varepsilon} \leq \bar\alpha, \quad R^6_{T-\varepsilon} \leq \bar\alpha + \frac{1}{2}\int_0^{T-\varepsilon} |\sigma^{-1}(t,\boldsymbol{\chi}_t)v_t - \sigma^{-1}(t,\boldsymbol{\phi}_t)\gamma_t|^2 dt. \tag{4.23}
$$

We choose $\bar\alpha = \ln(8)$ so that $1 - 2\exp(-\bar\alpha) = 3/4$ (i.e. $\mathbb{P}(\bar{\mathcal{A}}) \geq 3/4$). We then apply Proposition 4.3 with $\mu = 1/4$. For $T \leq c_{4.3}(1/4)$, there exists a constant C (only depending on (**A**)) and an event $\bar{\mathcal{B}}$, with $\mathbb{P}(\bar{\mathcal{B}}) \geq 3/4$, on which recalling that $\boldsymbol{\Theta}_t = \boldsymbol{\chi}_t - \boldsymbol{\phi}_t$ (see (4.2))

$$
\forall t \in [0, T-\varepsilon], \begin{cases} (T-t)|\mathbb{T}^{-1}_{T-t}(\boldsymbol{\chi}_t - \boldsymbol{\phi}_t - \boldsymbol{\Gamma}_t)| \leq C(T-t)^{\frac{1}{2}+\frac{\eta}{8}}, \\ |v_t^0 - \gamma_t| \leq C(T-t)^{-\frac{1}{2}+\frac{\eta}{8}}. \end{cases} \tag{4.24}
$$

Set now $\bar{\mathcal{C}} = \bar{\mathcal{A}} \cap \bar{\mathcal{B}}$, so that $\mathbb{P}(\bar{\mathcal{C}}) \geq 1/2$. By (4.23), $R^5_{T-\varepsilon} \leq C$ on $\bar{\mathcal{C}}$. By (4.23) again, (4.24), (1) in Proposition 4.2 and the inequality $|\mathbf{x}| \leq (T-t)|\mathbb{T}^{-1}_{T-t}\mathbf{x}|$ ($T \leq 1$), it also holds on $\bar{\mathcal{C}}$

$$
\begin{aligned}
R^6_{T-\varepsilon} &\leq C\Big[1 + \int_0^{T-\varepsilon} \Big(|\varphi_t|^2 + |v_t^0 - \gamma_t|^2 + |\boldsymbol{\chi}_t - \boldsymbol{\phi}_t|^{2\eta}|\gamma_t|^2\Big)dt\Big],\\
&\leq C + CT|\mathbb{T}_T^{-1}(\boldsymbol{\theta}_T(\mathbf{x}_0) - \mathbf{y}_0)|^2 + C\int_0^{T-\varepsilon} (T-t)^{\eta + \frac{\eta^2}{4}}|\gamma_t|^2 dt\\
&\quad + C\int_0^{T-\varepsilon} |\boldsymbol{\Gamma}_t|^{2\eta}|\gamma_t|^2 dt.
\end{aligned}
\tag{4.25}
$$

Once again, C is independent of T since $T \le 1$. Similarly, on the event $\bar{\mathcal{C}}$

$$
\begin{aligned}
R^2_{T-\varepsilon} &\le C \int_0^{T-\varepsilon} \langle a^{-1}(t, \boldsymbol{\phi}_t)[v_t^0 - \gamma_t], v_t^0 - \gamma_t + 2\gamma_t \rangle dt \\
&\le C\Big[1 + \int_0^{T-\varepsilon} (T-t)^{-\frac{1}{2}+\frac{\eta}{8}} |\gamma_t| dt\Big], \\
R^3_{T-\varepsilon} &\le C |\mathbb{T}_T^{-1}(\boldsymbol{\theta}_T(\mathbf{x}_0) - \mathbf{y}_0)| \Big[T^{1/2} + \int_0^{T-\varepsilon} |\gamma_t| dt\Big], \\
R^4_{T-\varepsilon} &\le C\Big[1 + \int_0^{T-\varepsilon} \Big(|\boldsymbol{\Gamma}_t|^\eta + (T-t)^{\frac{\eta}{2}+\frac{\eta^2}{8}}\Big)|\gamma_t|^2 dt + \int_0^{T-\varepsilon} |\boldsymbol{\Gamma}_t|^\eta (T-t)^{-1+\frac{\eta}{4}} dt\Big].
\end{aligned}
\tag{4.26}
$$

By Lemma 3.9, we know that $\mathbb{E}[|\boldsymbol{\Gamma}_t|^p] \le C(T-t)^{p/2}$ and $\mathbb{E}[|\gamma_t|^p] \le C(T-t)^{-p/2}, p \ge 1$. By (4.25) and (4.26) and Hölder's inequality, we obtain

$$
\mathbb{E}\Big[\mathbb{I}_{\bar{\mathcal{C}}}\Big(\sum_{i=2}^{6} R^i_{T-\varepsilon}\Big)\Big] \le C\Big(1 + T|\mathbb{T}_T^{-1}(\boldsymbol{\theta}_T(\mathbf{x}_0) - \mathbf{y}_0)|^2\Big). \tag{4.27}
$$

It finally remains to treat $R^1_{T-\varepsilon}$. As in the proof of Proposition 3.7, we denote by $\mathbf{K}(T-\varepsilon, T)$ the covariance matrix associated with $q(T-\varepsilon, T, \cdot, \cdot)$. Since $\eta_\varepsilon(\cdot) = q(T-\varepsilon, T, \cdot - \boldsymbol{\phi}_{T-\varepsilon}, \mathbf{0})$, we have from (3.2)

$$
\begin{aligned}
&|R^1_{T-\varepsilon}| \\
&\le C|\langle \mathbf{K}^{-1}(T-\varepsilon, T)\mathbf{R}(T, T-\varepsilon)[\boldsymbol{\chi}_{T-\varepsilon} - \boldsymbol{\phi}_{T-\varepsilon}], \mathbf{R}(T, T-\varepsilon)(\boldsymbol{\chi}_{T-\varepsilon} - \boldsymbol{\phi}_{T-\varepsilon})\rangle \\
&\quad - \langle \mathbf{K}^{-1}(T-\varepsilon, T)\mathbf{R}(T, T-\varepsilon)\boldsymbol{\Gamma}_{T-\varepsilon}, \mathbf{R}(T, T-\varepsilon)\boldsymbol{\Gamma}_{T-\varepsilon}\rangle| \\
&= C|\langle \mathbf{K}^{-1}(T-\varepsilon, T)\mathbf{R}(T, T-\varepsilon)[\boldsymbol{\chi}_{T-\varepsilon} - \boldsymbol{\phi}_{T-\varepsilon} - \boldsymbol{\Gamma}_{T-\varepsilon}], \\
&\qquad \mathbf{R}(T, T-\varepsilon)[\boldsymbol{\chi}_{T-\varepsilon} - \boldsymbol{\phi}_{T-\varepsilon} + \boldsymbol{\Gamma}_{T-\varepsilon}]\rangle|.
\end{aligned}
$$

(Here, $(\mathbf{R}(t,s))_{0 \le t,s \le T}$ is the resolvent associated with $(\mathbf{L}_t = D\boldsymbol{\mathcal{F}}(t, \mathbf{0}))_{0 \le t \le T}$.) By scaling Lemma 3.6, the matrix $[\mathbf{R}(T, T-\varepsilon)]^* \mathbf{K}^{-1}(T-\varepsilon, T)\mathbf{R}(T, T-\varepsilon)$ is less than $C\varepsilon \mathbb{T}_\varepsilon^{-2}$. (As required, C is independent of T since $T \le 1$.) By (4.24),

$$
\begin{aligned}
|R^1_{T-\varepsilon}| &\le C\varepsilon |\mathbb{T}_\varepsilon^{-1}(\boldsymbol{\chi}_{T-\varepsilon} - \boldsymbol{\phi}_{T-\varepsilon} - \boldsymbol{\Gamma}_{T-\varepsilon})||\mathbb{T}_\varepsilon^{-1}(\boldsymbol{\chi}_{T-\varepsilon} - \boldsymbol{\phi}_{T-\varepsilon} + \boldsymbol{\Gamma}_{T-\varepsilon})| \\
&\le C\varepsilon^{\frac{\eta}{8}}\Big(\varepsilon^{\frac{\eta}{8}} + \varepsilon^{\frac{1}{2}}|\mathbb{T}_\varepsilon^{-1}\boldsymbol{\Gamma}_{T-\varepsilon}|\Big).
\end{aligned}
$$

By Lemma 3.9,

$$
\mathbb{E}\Big[\mathbb{I}_{\bar{\mathcal{C}}} R^1_{T-\varepsilon}\Big] \le C\varepsilon^{\frac{\eta}{8}} \le C. \tag{4.28}
$$

Taking the expectation on $\bar{\mathcal{C}}$ in (4.21), we finally deduce from (4.27) and (4.28)

$$
J_\varepsilon(0, \mathbf{x}_0)\mathbb{P}(\bar{\mathcal{C}}) \le -\ln(q(T, 0, 0))\mathbb{P}(\bar{\mathcal{C}}) + C\Big(1 + T|\mathbb{T}_T^{-1}(\boldsymbol{\theta}_T(\mathbf{x}_0) - \mathbf{y}_0)|^2\Big).
$$

Since $\mathbb{P}(\bar{\mathcal{C}}) \ge 1/2$, we have

$$
J_\varepsilon(0, \mathbf{x}_0) \le \frac{1}{2}\ln\Big(\det(\mathbf{K}(0, T))\Big) + C\Big(1 + T|\mathbb{T}_T^{-1}(\boldsymbol{\theta}_T(\mathbf{x}_0) - \mathbf{y}_0)|^2\Big).
$$

Letting ε tend to 0, we deduce from Proposition 3.4 that, for $T \le c_{4.3}(1/4) \wedge 1$,

$$-\ln(p(T,\mathbf{x}_0,\mathbf{y}_0)) \le \frac{n^2 d}{2}\ln(T) + C\Big(1 + T|\mathbb{T}_T^{-1}(\boldsymbol{\theta}_T(\mathbf{x}_0) - \mathbf{y}_0)|^2\Big). \qquad (4.29)$$

Third Step. We emphasize that the constant C in (4.29) is independent of T. In particular, for a given $T \le c_{4.3}(1/4) \wedge 1$, (4.29) also holds for any $t \in (0,T]$. This proves the lower bound in Theorem 1.1 for $T \le c_{4.3}(1/4) \wedge 1$.

4.4 Scaling Argument

Using the intrinsic scaling properties of the system, see Subsection 2.3, Eq. (1.1) set on an interval $[0,T]$ of length $T > c_{4.3}(1/4) \wedge 1$ can be rescaled on the interval $[0, c_{4.3}(1/4) \wedge 1]$ up to a magnification of the constants in (**A**) by some power of T. Following Subsection 2.3, we can derive the expected lower bound for the density over an interval of arbitrary length from the short time estimate. $\square$

5 Upper Bound for the Density

We now prove the upper bound in Theorem 1.1. By the intrinsic scaling properties of the system, it is sufficient to prove it on $[0,1]$: an argument similar to the one used in Subsection 4.4, see also Section 2.3, permits to reduce the problem from $[0,T]$ to $[0,1]$ for any $T \ge 1$. In fact, it is even sufficient to prove the upper bound in Theorem 1.1 at time $t = 1$ only. Indeed, the scaling property (2.11) permits to write the density kernel $p(t,\cdot,\cdot)$ at time $t \le 1$ as $t^{-n^2 d/2}\hat{p}(1, t^{1/2}\mathbb{T}_t^{-1}\cdot, t^{1/2}\mathbb{T}_t^{-1}\cdot)$, for some rescaled density kernel $\hat{p}(1,\cdot,\cdot)$ at time 1 satisfying Assumption (**A**) with respect to the same constants κ, Λ and η as the original kernel p. We emphasize that we didn't use this latter argument to prove the lower bound: the method used in Subsection 4.3 directly provides the required estimate on the whole $[0, c_{4.3}(1/4) \wedge 1]$.

Throughout the section, we follow the strategy announced in Subsection 2.4. It relies on the McKean–Singer expansion (2.17) (with $T = 1$) that we here recall:

$$\begin{aligned} p(1,\mathbf{x}_0,\mathbf{y}_0) &= \tilde{p}(0,1,\mathbf{x}_0,\mathbf{y}_0) + \sum_{k=1}^{N}\int_0^1\int_{\mathbb{R}^{nd}} \tilde{p}(0,t,\mathbf{x}_0,\mathbf{z})H^{\otimes k}(t,1,\mathbf{z},\mathbf{y}_0)dtd\mathbf{z} \\ &\quad + \int_0^1\int_{\mathbb{R}^{nd}} p(t,\mathbf{x}_0,\mathbf{z})H^{\otimes(N+1)}(t,1,\mathbf{z},\mathbf{y}_0)dtd\mathbf{z}. \end{aligned} \qquad (5.1)$$

We remind the reader that, for any $0 < T \leq 1$, $\tilde{p}(0,T,\mathbf{x},\mathbf{y})$ stands for $\tilde{p}^{T,\mathbf{y}}(0,T,\mathbf{x},\mathbf{y})$, where $(\tilde{p}^{T,\mathbf{y}}(s,t,\mathbf{x},\mathbf{z}))_{0\leq s<t\leq T;\mathbf{x},\mathbf{z}\in\mathbb{R}^{nd}}$ is the transition density of some Markovian Gaussian process, obtained by linearization of (1.1). The kernel H together with its iterated products $(H^{\otimes k})_{k\geq 1}$ are associated with $\tilde{p}$ by formulas (2.16) and (2.18).

The Gaussian process we here consider for the construction of $\tilde{p}$ is a bit different from the announced version (2.15). Following Subsection 4.2, we indeed replace the "complete" gradient $\mathbf{D_x F}$ in (2.15) by its reduced version $D\mathbf{F}$. (That is we just consider the subdiagonal of $\mathbf{D_x F}$.) In what follows, for any $\mathbf{y} \in \mathbb{R}^{nd}$, $(\tilde{p}^{T,\mathbf{y}}(s,t,\mathbf{x},\mathbf{z}))_{0\leq s<t\leq T;\mathbf{x},\mathbf{z}\in\mathbb{R}^{nd}}$ thus denotes the transition density associated with the linear equation:

$$\begin{aligned} d\tilde{\mathbf{X}}_t = &\Big[\mathbf{F}(t,\boldsymbol{\theta}_{t,T}(\mathbf{y})) + D\mathbf{F}(t,\boldsymbol{\theta}_{t,T}(\mathbf{y}))\Big(\tilde{\mathbf{X}}_t - \boldsymbol{\theta}_{t,T}(\mathbf{y})\Big)\Big]dt \\ &+ B\sigma(t,\boldsymbol{\theta}_{t,T}(\mathbf{y}))dW_t, \quad 0 \leq t \leq T, \end{aligned} \tag{5.2}$$

where $(\boldsymbol{\theta}_{t,T}(\mathbf{y}))_{t\geq 0}$ solves the ODE $[d/dt]\boldsymbol{\theta}_{t,T}(\mathbf{y}) = \mathbf{F}(t,\boldsymbol{\theta}_{t,T}(\mathbf{y}))$, $t \geq 0$, with the boundary condition $\boldsymbol{\theta}_{T,T}(\mathbf{y}) = \mathbf{y}$. (See (2.14).) To simplify, we do not specify the dependence of $\tilde{\mathbf{X}}$ on $(T,\mathbf{y})$. (*In what follows, we will consider the flow for various initializing times, i.e. we will consider $(\boldsymbol{\theta}_{t,s})_{t\geq 0}$, $s \geq 0$, solution of the ODE driven by $(\mathbf{F}(t,\cdot))_{t\geq 0}$ with the boundary condition $\boldsymbol{\theta}_{s,s} = \mathrm{identity}$. We will also make use of the semigroup property $\boldsymbol{\theta}_{t,s} \circ \boldsymbol{\theta}_{s,r} = \boldsymbol{\theta}_{t,r}$.*)

The proof is divided into two propositions. The first one permits to estimate $\tilde{p}$ and H. The second one permits to handle the convolution of p and the iterated products of H. To state these propositions, we introduce the useful notation:

$$g_{a,t}(\mathbf{y}) = t^{-n^2\frac{d}{2}} \exp(-a^{-1}t|\mathbb{T}_t^{-1}\mathbf{y}|^2), \quad a,t > 0, \ \mathbf{y} \in \mathbb{R}^{nd}. \tag{5.3}$$

Up to a normalizing constant depending on a, n, d, $g_{a,t}$ is a Gaussian density. We emphasize that $(g_{a,t})_{a>0,t>0}$ satisfies Lemma B.1 in Appendix: we use it right below. In the following, $\boldsymbol{\theta}_t(\mathbf{x})$ is a short version for $\boldsymbol{\theta}_{t,0}(\mathbf{x})$.

Proposition 5.1 *There exists a family of constants $(C_{5.1}(N))_{N\geq 0}$, only depending on ($\mathbf{A}$), such that, for all $N \geq 1$, $0 < t < 1$, and $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{R}^{nd}$,*

$$\begin{aligned} &\tilde{p}(0,t,\mathbf{x},\mathbf{y}) \leq C_{5.1}(0) g_{C_{5.1}(0),t}\Big(\boldsymbol{\theta}_t(\mathbf{x}) - \mathbf{y}\Big), \\ &|H^{\otimes N}(t,1,\mathbf{z},\mathbf{y})| \leq C_{5.1}(N)(1-t)^{N\frac{\eta}{2}-1} g_{C_{5.1}(N),1-t}\Big(\mathbf{z} - \boldsymbol{\theta}_{t,1}(\mathbf{y})\Big). \end{aligned}$$

Proposition 5.2 *Let $a > 0$. Then, there exists a constant $C_{5.2}(a) > 0$, only depending on a and ($\mathbf{A}$), such that, for all $t \in [0,1)$ and $\mathbf{x}, \mathbf{y} \in \mathbb{R}^{nd}$,*

$$\int_{\mathbb{R}^{nd}} p(t,\mathbf{x},\mathbf{z})(1-t)^{\frac{n^2 d}{2}} g_{a,1-t}\Big(\mathbf{z} - \boldsymbol{\theta}_{t,1}(\mathbf{y})\Big) d\mathbf{z} \leq C_{5.2}(a) g_{C_{5.2}(a),1}\Big(\boldsymbol{\theta}_1(\mathbf{x}) - \mathbf{y}\Big).$$

The second equation in Proposition 5.1 gives the regularizing effect of the kernel in function of the regularity in **(A)**. This feature is characteristic of the parametrix techniques, see e.g. [MS67], [KM00]. Proposition 5.2 is crucial and allows to truncate the parametrix series expansion. It is the key point in order to handle the non-linearity of the degenerate terms in the operator $\mathcal{L}_t$ introduced after (2.1). This non-linearity breaks the underlying semigroup structure of Kolmogorov's example, see also [KMM09] and makes the truncation unavoidable.

5.1 From Propositions 5.1 and 5.2 to the Upper Bound

Admitting these two propositions, we derive the upper bound in Theorem 1.1 at time $t = 1$.

We have to show that $p(1, \mathbf{x}_0, \mathbf{y}_0)$ is less than $Cg_{C,1}(\boldsymbol{\theta}_1(\mathbf{x}_0) - \mathbf{y}_0)$ for some $C > 0$ only depending on (**A**). The proof relies on (5.1). The first term in (5.1), i.e. the Gaussian density, is easily bounded by using the first line in Proposition 5.1 with $t = 1$, $\mathbf{x} = \mathbf{x}_0$ and $\mathbf{y} = \mathbf{y}_0$. The last term in (5.1) is bounded by plugging the second line in Proposition 5.1 with $N + 1 = \lceil (n^2 d + 2)/\eta \rceil$ and $\mathbf{y} = \mathbf{y}_0$ into the bound in Proposition 5.2 with $\mathbf{x} = \mathbf{x}_0$ and $\mathbf{y} = \mathbf{y}_0$: indeed, $\lceil (n^2 d + 2)/\eta \rceil \eta/2 - 1 \geq (n^2 d + 2)/2 - 1 = n^2 d/2$. To handle the sum in (5.1), we apply Proposition 5.1 and Lemma B.1: we obtain, for $k \geq 1$,

$$
\begin{aligned}
&\int_0^1 \int_{\mathbb{R}^{nd}} \tilde{p}(0, t, \mathbf{x}_0, \mathbf{z}) |H^{\otimes k}(t, 1, \mathbf{z}, \mathbf{y}_0)| d\mathbf{z} dt \\
&\leq C(k) \int_0^1 \int_{\mathbb{R}^{nd}} (1-t)^{k\frac{\eta}{2}-1} g_{C(k),t}\Big(\boldsymbol{\theta}_t(\mathbf{x}_0) - \mathbf{z}\Big) g_{C(k),1-t}\Big(\mathbf{z} - \boldsymbol{\theta}_{t,1}(\mathbf{y}_0)\Big) d\mathbf{z} dt \\
&\leq C(k) \int_0^1 (1-t)^{k\frac{\eta}{2}-1} g_{C(k),1}\Big(\boldsymbol{\theta}_t(\mathbf{x}_0) - \boldsymbol{\theta}_{t,1}(\mathbf{y}_0)\Big) dt,
\end{aligned}
\tag{5.4}
$$

the constant $C(k)$ only depending on k and (**A**). By the semi-group property of the flow, $|\boldsymbol{\theta}_t(\mathbf{x}_0) - \boldsymbol{\theta}_{t,1}(\mathbf{y}_0)| = |\boldsymbol{\theta}_{t,1}(\boldsymbol{\theta}_1(\mathbf{x}_0)) - \boldsymbol{\theta}_{t,1}(\mathbf{y}_0)|$. Similarly, $\boldsymbol{\theta}_{t,1}$ is a diffeomorphic mapping from $\mathbb{R}^{nd}$ onto itself with $\boldsymbol{\theta}_{1,t}$ as converse. In particular, $\boldsymbol{\theta}_{t,1}$ is a Lipschitz diffeomorphism so that $|\boldsymbol{\theta}_t(\mathbf{x}_0) - \boldsymbol{\theta}_{t,1}(\mathbf{y}_0)| = |\boldsymbol{\theta}_{t,1}(\boldsymbol{\theta}_1(\mathbf{x}_0)) - \boldsymbol{\theta}_{t,1}(\mathbf{y}_0)| \geq C^{-1}|\boldsymbol{\theta}_1(\mathbf{x}_0) - \mathbf{y}_0|$ for some constant C depending on (**A**) only. (C may be chosen independently of t since the Lipschtiz constant of $\boldsymbol{\theta}_{1,t}$ is uniform in $0 < t < 1$.) Since $g_{C(k),1}(\mathbf{y}) \leq g_{C(k),1}(\mathbf{y}')$ for $|\mathbf{y}| \geq |\mathbf{y}'|$, this completes the proof. □

5.2 Proof of Proposition 5.1

For $\mathbf{y} \in \mathbb{R}^{nd}$ and for $0 < T \leq 1$, we first give the form of the kernel $(\tilde{p}^{T,\mathbf{y}}(t, T, \mathbf{x}, \mathbf{z}))_{0 \leq t < T, \mathbf{x}, \mathbf{z} \in \mathbb{R}^{nd}}$ associated with the Gaussian process $\tilde{\mathbf{X}}$ in (5.2).

The deterministic ODE associated with $\tilde{\mathbf{X}}$ has the form

$$\frac{d}{dt}\tilde{\boldsymbol{\phi}}_t = \mathbf{F}(t, \boldsymbol{\theta}_{t,T}(\mathbf{y})) + D\mathbf{F}(t, \boldsymbol{\theta}_{t,T}(\mathbf{y}))[\tilde{\boldsymbol{\phi}}_t - \boldsymbol{\theta}_{t,T}(\mathbf{y})], \quad t \geq 0. \tag{5.5}$$

We denote by $(\tilde{\boldsymbol{\theta}}_{t,s}^{T,\mathbf{y}})_{s,t\geq 0}$ the associated flow, i.e. $\tilde{\boldsymbol{\theta}}_{t,s}^{T,\mathbf{y}}(\mathbf{x})$ is the value of $\tilde{\boldsymbol{\phi}}_t$ when $\tilde{\boldsymbol{\phi}}_s = \mathbf{x}$. It is affine:

$$\begin{aligned}\tilde{\boldsymbol{\theta}}_{t,s}^{T,\mathbf{y}}(\mathbf{x}) &= \tilde{\mathbf{R}}^{T,\mathbf{y}}(t,s)\mathbf{x} \\ &\quad + \int_s^t \tilde{\mathbf{R}}^{T,\mathbf{y}}(t,u)\Big(\mathbf{F}(u, \boldsymbol{\theta}_{u,T}(\mathbf{y})) - D\mathbf{F}(u, \boldsymbol{\theta}_{u,T}(\mathbf{y}))\boldsymbol{\theta}_{u,T}(\mathbf{y})\Big)du.\end{aligned} \tag{5.6}$$

Above, $(\tilde{\mathbf{R}}^{T,\mathbf{y}}(t,s))_{s,t\geq 0}$ stands for the resolvent associated with the matrices $(D\mathbf{F}(t, \boldsymbol{\theta}_{t,T}(\mathbf{y})))_{t\geq 0}$. We then claim

Lemma 5.3 *There exists a constant $C_{5.3} \geq 1$, depending on ($\mathbf{A}$) only (an not on T), such that, for any $t \in [0,T)$ and $\mathbf{x}, \mathbf{y} \in \mathbb{R}^{nd}$,*

$$C_{5.3}^{-1}\Big|\mathbb{T}_{T-t}^{-1}\Big[\mathbf{x} - \boldsymbol{\theta}_{t,T}(\mathbf{y})\Big]\Big| \leq \Big|\mathbb{T}_{T-t}^{-1}\Big[\tilde{\boldsymbol{\theta}}_{T,t}^{T,\mathbf{y}}(\mathbf{x}) - \mathbf{y}\Big]\Big| \leq C_{5.3}\Big|\mathbb{T}_{T-t}^{-1}\Big[\mathbf{x} - \boldsymbol{\theta}_{t,T}(\mathbf{y})\Big]\Big|.$$

Proof. It is clear that $\tilde{\boldsymbol{\theta}}_{t,T}^{T,\mathbf{y}}(\mathbf{y}) = \boldsymbol{\theta}_{t,T}(\mathbf{y})$ for $0 \leq t \leq T$. Indeed, $(\boldsymbol{\theta}_{t,T}(\mathbf{y}))_{0\leq t\leq T}$ satisfies (5.5) and matches $\mathbf{y}$ at time T. We thus compare $|\mathbb{T}_{T-t}^{-1}[\mathbf{x} - \tilde{\boldsymbol{\theta}}_{t,T}^{T,\mathbf{y}}(\mathbf{y})]|$ and $|\mathbb{T}_{T-t}^{-1}[\tilde{\boldsymbol{\theta}}_{T,t}^{T,\mathbf{y}}(\mathbf{x}) - \mathbf{y}]|$. By a translation argument, we can assume $t = 0$. The result then follows from (2.13) (applied to the linearized equation (5.5)). □

We deduce:

Lemma 5.4 *There exists a constant $C_{5.4} > 0$, depending on ($\mathbf{A}$) only (and not on T), such that, for all $0 \leq t < T$ and $\mathbf{x}, \mathbf{y} \in \mathbb{R}^{nd}$,*

$$\tilde{p}(t, T, \mathbf{x}, \mathbf{y}) \leq C_{5.4} g_{C_{5.4}, T-t}\Big(\mathbf{x} - \boldsymbol{\theta}_{t,T}(\mathbf{y})\Big).$$

Proof. The proof is almost direct. When initialized at point $\mathbf{x}$ at time t, the process $\tilde{\mathbf{X}}$ in (5.2) is a Gaussian process. Denoting by $\tilde{\mathbf{m}}(t,T)$ the mean and by $\tilde{\mathbf{K}}(t,T)$ the covariance matrix of the random vector $\tilde{\mathbf{X}}_T$, we know that

$$\tilde{p}(t, T, \mathbf{x}, \mathbf{y}) = (2\pi)^{-\frac{nd}{2}} \det{}^{-\frac{1}{2}}(\tilde{\mathbf{K}}(t,T)) \exp\Big(-\frac{|\tilde{\mathbf{K}}^{-\frac{1}{2}}(t,T)(\mathbf{y} - \tilde{\mathbf{m}}(t,T))|^2}{2}\Big). \tag{5.7}$$

It is plain to see that the mean of the Gaussian process $\tilde{\mathbf{X}}$ satisfies the ODE (5.5), so that $\tilde{\mathbf{m}}(t,T) = \tilde{\boldsymbol{\theta}}_{T,t}^{T,\mathbf{y}}(\mathbf{x})$. From Lemma 3.6, see also the proof of Proposition 3.7, we know that $\tilde{\mathbf{K}}(t,T)$ has the form $(T-t)^{-1}\mathbb{T}_{T-t}\hat{\mathbf{K}}_1^{t,T}\mathbb{T}_{T-t}$, where $\hat{\mathbf{K}}_1^{t,T}$ is the covariance matrix associated with some rescaled version of $\tilde{\mathbf{X}}$. (We do not specify what the rescaled version is.) Under ($\mathbf{A}$), the eigenvalues of the matrix $\hat{\mathbf{K}}_1^{t,T}$ are bounded from above and from below by known parameters uniformly in $0 \le t < T \le 1$, see Proposition 3.4. (In particular, the bounds are independent of t, T.) We deduce the expected bounds for $\tilde{p}(t,T,\mathbf{x},\mathbf{y})$ but with $\tilde{\boldsymbol{\theta}}_{T,t}^{T,\mathbf{y}}(\mathbf{x}) - \mathbf{y}$ instead of $\mathbf{x} - \boldsymbol{\theta}_{t,T}(\mathbf{y})$. By Lemma 5.3, we complete the proof. □

We now provide a first estimate for H. (Have in mind that $T \le 1$.)

Lemma 5.5 *There exists a constant $C_{5.5} > 0$, depending on* ($\mathbf{A}$) *only (and not on T), such that, for all $t \in [0,T)$ and $\mathbf{z}, \mathbf{y} \in \mathbb{R}^{nd}$,*

$$|H(t,T,\mathbf{z},\mathbf{y})| \le C_{5.5}(T-t)^{\frac{\eta}{2}-1} g_{C_{5.5},T-t}\Big(\mathbf{z} - \boldsymbol{\theta}_{t,T}(\mathbf{y})\Big).$$

Proof. The definition of H is given in (2.16). For a smooth function $f : \mathbf{z} \in \mathbb{R}^{nd} \mapsto f(\mathbf{z})$, we have, for $t \in [0,T)$ and $\mathbf{z} \in \mathbb{R}^{nd}$,

$$\begin{aligned}
&\Big[\mathcal{L}_{t,\mathbf{z}} f - \tilde{\mathcal{L}}_{t,\mathbf{z}}^{T,\mathbf{y}} f\Big](\mathbf{z}) = (1/2)\mathrm{Tr}\Big[\Big(a(t,\mathbf{z}) - a(t,\boldsymbol{\theta}_{t,T}(\mathbf{y}))\Big) D_{\mathbf{z}_1}^2 f(\mathbf{z})\Big] \\
&+ \sum_{j=1}^{n} \langle \mathbf{F}_j(t,\mathbf{z}) - \mathbf{F}_j(t,\boldsymbol{\theta}_{t,T}(\mathbf{y})) - D_{\mathbf{z}_{j-1}}\mathbf{F}_j(t,\boldsymbol{\theta}_{t,T}(\mathbf{y}))(\mathbf{z} - \boldsymbol{\theta}_{t,T}(\mathbf{y}))_{j-1}, D_{\mathbf{z}_j} f(\mathbf{z})\rangle,
\end{aligned}$$

with the convention that $D_{\mathbf{z}_0}\mathbf{F}_1 = 0$. From the specific structure of $\mathbf{F}$, we can find a constant $C > 0$, depending on ($\mathbf{A}$) only, such that

$$\begin{aligned}
\Big|\mathcal{L}_{t,\mathbf{z}} f(\mathbf{z}) - \tilde{\mathcal{L}}_{t,\mathbf{z}}^{T,\mathbf{y}} f(\mathbf{z})\Big| &\le C|\mathbf{z} - \boldsymbol{\theta}_{t,T}(\mathbf{y})|^{\eta} |D_{\mathbf{z}_1}^2 f(\mathbf{z})| \\
&\quad + C\sum_{j=2}^{n} |\mathbf{z}_{j-1} - (\boldsymbol{\theta}_{t,T}(\mathbf{y}))_{j-1}|^{1+\eta} |D_{\mathbf{z}_j} f(\mathbf{z})| \\
&\quad + C\sum_{j=1}^{n}\sum_{k=j}^{n} |\mathbf{z}_k - (\boldsymbol{\theta}_{t,T}(\mathbf{y}))_k||D_{\mathbf{z}_j} f(\mathbf{z})|.
\end{aligned} \tag{5.8}$$

We now apply the above inequality with $f(\mathbf{z}) = \tilde{p}(t,T,\mathbf{z},\mathbf{y})$, i.e. (see (5.7))

$$f(\mathbf{z}) = (2\pi)^{-\frac{nd}{2}} \det{}^{-\frac{1}{2}}(\tilde{\mathbf{K}}(t,T)) \exp\Big(-\frac{|\tilde{\mathbf{K}}^{-\frac{1}{2}}(t,T)(\mathbf{y} - \tilde{\boldsymbol{\theta}}_{T,t}(\mathbf{z}))|^2}{2}\Big).$$

(For simplicity, we will omit the superscript $(T,\mathbf{y})$ in $\tilde{\boldsymbol{\theta}}$ and $\tilde{\mathcal{L}}$.) We now expand the derivatives of f. To this end, we recall that $\tilde{\boldsymbol{\theta}}_{T,t}$ is an affine transformation with $\tilde{\mathbf{R}}(T,t)$ as linear part, see (5.6). (We will also omit the superscript $(T,\mathbf{y})$ in $\tilde{\mathbf{R}}$.)

$$\begin{aligned}
D_{\mathbf{z}_j} f(\mathbf{z}) &= -\Big[[\tilde{\mathbf{R}}(T,t)]^* \tilde{\mathbf{K}}^{-1}(t,T)(\tilde{\boldsymbol{\theta}}_{T,t}(\mathbf{z}) - \mathbf{y})\Big]_j \tilde{p}(t,T,\mathbf{z},\mathbf{y}), \\
D_{\mathbf{z}_j}^2 f(\mathbf{z}) &= -\Big[[\tilde{\mathbf{R}}(T,t)]^* \tilde{\mathbf{K}}^{-1}(t,T)\tilde{\mathbf{R}}(T,t)\Big]_{j,j} \tilde{p}(t,T,\mathbf{z},\mathbf{y}) \\
&\quad + \Big[[\tilde{\mathbf{R}}(T,t)]^* \tilde{\mathbf{K}}^{-1}(t,T)(\tilde{\boldsymbol{\theta}}_{T,t}(\mathbf{z}) - \mathbf{y})\Big]_j^{\otimes 2} \tilde{p}(t,T,\mathbf{z},\mathbf{y}),
\end{aligned}$$

where $[[\tilde{\mathbf{R}}(T,t)]^*\tilde{\mathbf{K}}^{-1}(t,T)\tilde{\mathbf{R}}(T,t)]_{j,j}$ stands for the j^{th} diagonal block of size d of the $nd \times nd$-matrix $[\tilde{\mathbf{R}}(T,t)]^*\tilde{\mathbf{K}}^{-1}(t,T)\tilde{\mathbf{R}}(T,t)$. By (5.8),

$$
\begin{aligned}
&\Big|\Big(\mathcal{L}_{t,\mathbf{z}} - \tilde{\mathcal{L}}_{t,\mathbf{z}}\Big)\tilde{p}(t,T,\mathbf{z},\mathbf{y})\Big| \\
&\le C\tilde{p}(t,T,\mathbf{z},\mathbf{y}) \times \Big\{\Big|\mathbf{z} - \boldsymbol{\theta}_{t,T}(\mathbf{y})\Big|^\eta \Big|\Big[[\tilde{\mathbf{R}}(T,t)]^*\tilde{\mathbf{K}}^{-1}(t,T)\tilde{\mathbf{R}}(t,T)\Big]_{1,1}\Big| \\
&+ \ \Big|\mathbf{z} - \boldsymbol{\theta}_{t,T}(\mathbf{y})\Big|^\eta \Big|\Big[[\tilde{\mathbf{R}}(T,t)]^*\tilde{\mathbf{K}}^{-1}(t,T)(\mathbf{y} - \tilde{\boldsymbol{\theta}}_{T,t}(\mathbf{z}))\Big]_1\Big|^2\Big] \\
&+ \sum_{j=2}^n \Big|\mathbf{z}_{j-1} - (\boldsymbol{\theta}_{t,T}(\mathbf{y}))_{j-1}\Big|^{1+\eta}\Big|\Big[[\tilde{\mathbf{R}}(T,t)]^*\tilde{\mathbf{K}}^{-1}(t,T)\Big(\mathbf{y} - \tilde{\boldsymbol{\theta}}_{T,t}(\mathbf{z})\Big)\Big]_j\Big| \\
&+ \sum_{j=1}^n\sum_{k=j}^n \Big|\mathbf{z}_k - (\boldsymbol{\theta}_{t,T}(\mathbf{y}))_k\Big| \Big|\Big[[\tilde{\mathbf{R}}(T,t)]^*\tilde{\mathbf{K}}^{-1}(t,T)\Big(\mathbf{y} - \tilde{\boldsymbol{\theta}}_{T,t}(\mathbf{z})\Big)\Big]_j\Big|\Big\}.
\end{aligned}
\tag{5.9}
$$

We write $[\tilde{\mathbf{R}}(T,t)]^*\tilde{\mathbf{K}}^{-1}(t,T)$ as $(T-t)[\tilde{\mathbf{R}}(T,t)]^*\mathbb{T}_{T-t}^{-1}(\hat{\mathbf{K}}_1^{t,T})^{-1}\mathbb{T}_{T-t}^{-1}$, the matrix $\hat{\mathbf{K}}_1^{t,T}$ standing for the covariance matrix at time 1 of some rescaled process satisfying Assumption (**A**), see the proof of Lemma 5.4. By similar scaling arguments, see also the proof of Proposition 3.7, we know that $[\tilde{\mathbf{R}}(T,t)]^*\mathbb{T}_{T-t}^{-1} = \mathbb{T}_{T-t}^{-1}[\hat{\mathbf{R}}^{t,T}(1,0)]^*$, where $\hat{\mathbf{R}}^{t,T}(1,0)$ stands for the resolvent associated with the rescaled version of $\tilde{\mathbf{X}}$. We deduce that

$$
\begin{aligned}
&\Big[[\tilde{\mathbf{R}}(T,t)]^*\tilde{\mathbf{K}}^{-1}(t,T)\tilde{\mathbf{R}}(T,t)\Big]_{1,1} \\
&= (T-t)\Big[\mathbb{T}_{T-t}^{-1}[\hat{\mathbf{R}}^{t,T}(1,0)]^*(\hat{\mathbf{K}}_1^{t,T})^{-1}\hat{\mathbf{R}}^{t,T}(1,0)\mathbb{T}_{T-t}^{-1}\Big]_{1,1} \\
&= (T-t)^{-1}\Big[[\hat{\mathbf{R}}^{t,T}(1,0)]^*(\hat{\mathbf{K}}_1^{t,T})^{-1}\hat{\mathbf{R}}^{t,T}(1,0)\Big]_{1,1} \le C(T-t)^{-1}I_d
\end{aligned}
\tag{5.10}
$$

for some C, depending on (**A**) only (and not on T). (I_d the identity matrix of size d.) Similarly, for each $1 \le j \le d$,

$$
\begin{aligned}
&\Big|\Big[[\tilde{\mathbf{R}}(T,t)]^*\tilde{\mathbf{K}}^{-1}(t,T)(\mathbf{y} - \tilde{\boldsymbol{\theta}}_{T,t}(\mathbf{z}))\Big]_j\Big| \\
&= (T-t)^{-j+1}\Big|\Big[[\hat{\mathbf{R}}^{t,T}(1,0)]^*(\hat{\mathbf{K}}_1^{t,T})^{-1}\mathbb{T}_{T-t}^{-1}(\mathbf{y} - \tilde{\boldsymbol{\theta}}_{T,t}(\mathbf{z}))\Big]_j\Big| \\
&\le C(T-t)^{-j+1}\Big|\mathbb{T}_{T-t}^{-1}(\mathbf{y} - \tilde{\boldsymbol{\theta}}_{T,t}(\mathbf{z}))\Big|.
\end{aligned}
\tag{5.11}
$$

Plugging (5.10) and (5.11) into (5.9), we deduce

$$
\begin{aligned}
&\Big|\Big(\mathcal{L}_{\mathbf{z}} - \tilde{\mathcal{L}}_{\mathbf{z}}\Big)\tilde{p}(t,T,\mathbf{z},\mathbf{y})\Big| \\
&\le C\tilde{p}(t,T,\mathbf{z},\mathbf{y})\Big[|\mathbf{z} - \boldsymbol{\theta}_{t,T}(\mathbf{y})|^\eta\Big((T-t)^{-1} + |\mathbb{T}_{T-t}^{-1}(\mathbf{y} - \tilde{\boldsymbol{\theta}}_{T,t}(\mathbf{z}))|^2\Big) \\
&+ \sum_{j=2}^n (T-t)^{-j+1}\Big|\mathbf{z}_{j-1} - (\boldsymbol{\theta}_{t,T}(\mathbf{y}))_{j-1}\Big|^{1+\eta}\Big|\mathbb{T}_{T-t}^{-1}(\mathbf{y} - \tilde{\boldsymbol{\theta}}_{T,t}(\mathbf{z}))\Big| \\
&+ \sum_{j=1}^n\sum_{k=j}^n (T-t)^{-j+1}\Big|\mathbf{z}_k - (\boldsymbol{\theta}_{t,T}(\mathbf{y}))_k\Big|\Big|\mathbb{T}_{T-t}^{-1}(\mathbf{y} - \tilde{\boldsymbol{\theta}}_{T,t}(\mathbf{z}))\Big|\Big].
\end{aligned}
$$

Now, it is clear that $|\mathbf{z}_\ell-(\boldsymbol{\theta}_{t,T}(\mathbf{y}))_\ell| \le (T-t)^\ell|\mathbb{T}_{T-t}^{-1}(\mathbf{z}-\boldsymbol{\theta}_{t,T}(\mathbf{y}))|$. In particular, $|\mathbf{z}-\boldsymbol{\theta}_{t,T}(\mathbf{y})| \le (T-t)|\mathbb{T}_{T-t}^{-1}(\mathbf{z}-\boldsymbol{\theta}_{t,T}(\mathbf{y}))|$ since $T \le 1$. By Lemma 5.3 with $\mathbf{x}=\mathbf{z}$,

$$
\begin{aligned}
&\Big|\Big(\mathcal{L}_{t,\mathbf{z}}-\tilde{\mathcal{L}}_{t,\mathbf{z}}\Big)\tilde{p}(t,T,\mathbf{z},\mathbf{y})\Big| \\
&\le C\tilde{p}(t,T,\mathbf{z},\mathbf{y})\Big[(T-t)^{-1+\eta}|\mathbb{T}_{T-t}^{-1}(\mathbf{z}-\boldsymbol{\theta}_{t,T}(\mathbf{y}))|^\eta \\
&\qquad + (T-t)|\mathbb{T}_{T-t}^{-1}(\mathbf{z}-\boldsymbol{\theta}_{t,T}(\mathbf{y}))|^2 \\
&\qquad + (T-t)^\eta|\mathbb{T}_{T-t}^{-1}(\mathbf{z}-\boldsymbol{\theta}_{t,T}(\mathbf{y}))|^{2+\eta}\Big] \\
&\le C(T-t)^{-1+\frac{\eta}{2}}\tilde{p}(t,T,\mathbf{z},\mathbf{y})\Big[(T-t)^{\frac{\eta}{2}}|\mathbb{T}_{T-t}^{-1}(\mathbf{z}-\boldsymbol{\theta}_{t,T}(\mathbf{y}))|^\eta \\
&\qquad + (T-t)|\mathbb{T}_{T-t}^{-1}(\mathbf{z}-\boldsymbol{\theta}_{t,T}(\mathbf{y}))|^2 \\
&\qquad + (T-t)^{1+\frac{\eta}{2}}|\mathbb{T}_{T-t}^{-1}(\mathbf{z}-\boldsymbol{\theta}_{t,T}(\mathbf{y}))|^{2+\eta}\Big].
\end{aligned}
$$

By Lemma 5.4 (with $\mathbf{x}=\mathbf{z}$), there exists a constant C, only depending on $(\mathbf{A})$ (and not on T), such that,

$$\Big|\Big(\mathcal{L}_{t,\mathbf{z}}-\tilde{\mathcal{L}}_{t,\mathbf{z}}\Big)\tilde{p}(t,T,\mathbf{z},\mathbf{y})\Big| \le C(T-t)^{-1+\frac{\eta}{2}}g_{C,T-t}\Big(|\mathbb{T}_{1-t}^{-1}(\mathbf{z}-\boldsymbol{\theta}_{t,T}(\mathbf{y}))|^2\Big). \qquad \square$$

End of the Proof of Proposition 5.1. It remains to prove that, for every integer $N \ge 1$, there exists a constant $C(N)$, depending on N and $(\mathbf{A})$ only, such that, for any $0 \le t < 1$,

$$|H^{\otimes N}(t,1,\mathbf{z},\mathbf{y})| \le C(N)(1-t)^{N\frac{\eta}{2}-1}g_{C(N),1-t}\Big(\mathbf{z}-\boldsymbol{\theta}_{t,1}(\mathbf{y})\Big). \tag{5.12}$$

By Lemma 5.5, we know that (5.12) holds when $N=1$. We now perform an induction to complete the proof. We thus assume that (5.12) holds for some integer $N \ge 1$. Then, by Lemma 5.5,

$$
\begin{aligned}
&|H^{\otimes(N+1)}(t,1,\mathbf{z},\mathbf{y})| \\
&= \Big|\int_t^1\int_{\mathbb{R}^{nd}} H(t,s,\mathbf{z},\mathbf{z}')H^{\otimes N}(s,1,\mathbf{z}',\mathbf{y})dsd\mathbf{z}'\Big| \\
&\le C(N)C(1)\int_t^1\int_{\mathbb{R}^{nd}}\Big[(s-t)^{-1+\frac{\eta}{2}}(1-s)^{-1+N\frac{\eta}{2}} \\
&\qquad g_{C(1),s-t}\Big(\mathbf{z}-\boldsymbol{\theta}_{t,s}(\mathbf{z}')\Big)g_{C(N),1-s}\Big(\mathbf{z}'-\boldsymbol{\theta}_{s,1}(\mathbf{y})\Big)dsd\mathbf{z}'\Big].
\end{aligned}
$$

By (2.13), $C^{-1}|\mathbb{T}_{s-t}^{-1}(\mathbf{z}-\boldsymbol{\theta}_{t,s}(\mathbf{z}'))| \leq |\mathbb{T}_{s-t}^{-1}(\boldsymbol{\theta}_{s,t}(\mathbf{z})-\mathbf{z}')| \leq C|\mathbb{T}_{s-t}^{-1}(\mathbf{z}-\boldsymbol{\theta}_{t,s}(\mathbf{z}'))|$, for some constant C depending on $(\mathbf{A})$ only. Therefore, by Lemma B.1,

$$
\begin{aligned}
&|H^{\otimes(N+1)}(t,1,\mathbf{z},\mathbf{y})| \\
&\leq C(N)C(1)\int_t^1\int_{\mathbb{R}^{nd}}\Big[(s-t)^{-1+\frac{\eta}{2}}(1-s)^{-1+N\frac{\eta}{2}} \\
&\qquad\qquad g_{CC(1),s-t}\Big(\boldsymbol{\theta}_{s,t}(\mathbf{z})-\mathbf{z}'\Big)g_{C(N),1-s}\Big(\mathbf{z}'-\boldsymbol{\theta}_{s,1}(\mathbf{y})\Big)dsd\mathbf{z}'\Big] \\
&\leq C(N+1)\int_t^1(s-t)^{-1+\frac{\eta}{2}}(1-s)^{-1+N\frac{\eta}{2}}g_{C(N+1),1-t}\Big(\boldsymbol{\theta}_{s,t}(\mathbf{z})-\boldsymbol{\theta}_{s,1}(\mathbf{y})\Big)ds,
\end{aligned}
$$

for a constant $C(N+1)$, depending on the required parameters only. To complete the proof, it remains to see that $\int_t^1(s-t)^{-1+\eta/2}(1-s)^{-1+N\eta/2}ds = \beta(\eta/2,N\eta/2)(1-t)^{-1+(N+1)\eta/2}$ and that

$$
g_{C(N+1),1-t}\Big(\boldsymbol{\theta}_{s,t}(\mathbf{z})-\boldsymbol{\theta}_{s,1}(\mathbf{y})\Big) \leq g_{C'(N+1),1-t}\Big(\mathbf{z}-\boldsymbol{\theta}_{t,1}(\mathbf{y})\Big), \quad t\leq s\leq 1, \tag{5.13}
$$

for a constant $C'(N+1)$ depending on the same parameters as $C(N+1)$ (and thus independent of s). To obtain (5.13), it is sufficient to prove that, for any $s\in[t,1]$, $|\mathbb{T}_{1-t}^{-1}(\mathbf{z}-\boldsymbol{\theta}_{t,1}(\mathbf{y}))| \leq C|\mathbb{T}_{1-t}^{-1}(\boldsymbol{\theta}_{s,t}(\mathbf{z})-\boldsymbol{\theta}_{s,1}(\mathbf{y}))|$ for a constant C, depending on $(\mathbf{A})$ only (and not on s). The strategy is the same as in (2.12)–(2.13): we write $\mathbf{z}-\boldsymbol{\theta}_{t,1}(\mathbf{y}) = \boldsymbol{\theta}_{t,s}(\boldsymbol{\theta}_{s,t}(\mathbf{z}))-\boldsymbol{\theta}_{t,s}(\boldsymbol{\theta}_{s,1}(\mathbf{y})) = (1-t)^{-1/2}\mathbb{T}_{1-t}(\hat{\boldsymbol{\theta}}_{0,(s-t)/(1-t)}((1-t)^{1/2}\mathbb{T}_{1-t}^{-1}\boldsymbol{\theta}_{s,t}(\mathbf{z}))) - (1-t)^{-1/2}\mathbb{T}_{1-t}(\hat{\boldsymbol{\theta}}_{0,(s-t)/(1-t)}((1-t)^{1/2}\mathbb{T}_{1-t}^{-1}\boldsymbol{\theta}_{s,1}(\mathbf{y})))$, where $\hat{\boldsymbol{\theta}}$ stands for some bi-Lipschitz rescaled flow. Using the Lipschitz property of $\hat{\boldsymbol{\theta}}_{0,(s-t)/(1-t)}$, the proof is easily completed.

We emphasize that the modification of the constant $C(N)$ in (5.12) as N increases explains why we need to truncate the McKean-Singer expansion of the density. The modification follows from the transport of the initial condition by the flow $\boldsymbol{\theta}$. □

5.3 Proof of Proposition 5.2

The points $\mathbf{x}$ and $\mathbf{y}$ as well as the parameters a and t are fixed for the whole proof. For an arbitrary density q on $\mathbb{R}^{nd}$, we then set

$$
J^q := -\ln\int_{\mathbb{R}^{nd}}p(t,\mathbf{x},\mathbf{z})q(\mathbf{z})d\mathbf{z} = -\ln\Big(\mathbb{E}q(\mathbf{X}_t^{0,\mathbf{x}})\Big).
$$

We remind the reader of the representation formula (2.6):

$$
J^q = \inf_{(v_s)_s\in\mathcal{P}(t)}\mathbb{E}\Big[\frac{1}{2}\int_0^t\langle a^{-1}(s,\boldsymbol{\chi}_s)v_s,v_s\rangle ds - \ln\Big(q(\boldsymbol{\chi}_t)\Big)\Big], \tag{5.14}
$$

$(\boldsymbol{\chi}_s)_{0\le s\le t}$ standing for the controlled process

$$d\boldsymbol{\chi}_s = [\mathbf{F}(s,\boldsymbol{\chi}_s) + Bv_s]ds + B\sigma(s,\boldsymbol{\chi}_s)dW_s, \quad 0 \le s \le t,$$

with the initial condition $\boldsymbol{\chi}_0 = \mathbf{x}$. (See (2.5).) For a given control $(v_s)_{0\le s\le t}$, we set $\tilde{\boldsymbol{\chi}}_s = \boldsymbol{\chi}_s - \int_0^s B\sigma(u,\boldsymbol{\chi}_u)dW_u$, $0 \le s \le t$. It satisfies the controlled ODE (with random coefficients):

$$\begin{aligned} d\tilde{\boldsymbol{\chi}}_s &= [\mathbf{F}(s,\boldsymbol{\chi}_s) + Bv_s]ds \\ &= \Big[\mathbf{F}\Big(s, \int_0^s B\sigma(u,\boldsymbol{\chi}_u)dW_u + \tilde{\boldsymbol{\chi}}_s\Big) + Bv_s\Big]ds \\ &= [\mathbf{G}(s,\tilde{\boldsymbol{\chi}}_s) + Bv_s]ds, \end{aligned} \tag{5.15}$$

where $\mathbf{G}(s,\cdot) = \mathbf{F}(s, \int_0^s B\sigma(u,\boldsymbol{\chi}_u)dW_u + \cdot)$ is a progressively measurable process. For each random outcome ω, we emphasize that $(s,\mathbf{x}) \mapsto \mathbf{G}(s,\mathbf{x})$ satisfies (**A**): we can see (5.15) as a deterministic control problem of the form (2.8)–(2.9) (but with random coefficients). In particular, for each ω, the cost of the control $(v_s)_{0\le s\le t}$ is greater than the minimal cost to go from $\mathbf{x}$ to $\tilde{\boldsymbol{\chi}}_t$. By Proposition 4.1 and by the inequality $t|\mathbb{T}_t^{-1}\mathbf{z}|^2 \ge |\mathbf{z}|^2$, $\mathbf{z} \in \mathbb{R}^{nd}$, (which holds true since $t \le 1$), we obtain

$$\int_0^t |v_s|^2 dt \ge C_{4.1}^{-1} t |\mathbb{T}_t^{-1}(\tilde{\boldsymbol{\theta}}_t(\mathbf{x}) - \tilde{\boldsymbol{\chi}}_t)|^2 \ge C_{4.1}^{-1}|\tilde{\boldsymbol{\theta}}_t(\mathbf{x}) - \tilde{\boldsymbol{\chi}}_t|^2, \tag{5.16}$$

where $(\tilde{\boldsymbol{\theta}}_s(\mathbf{x}))_{0\le s\le t}$ is the solution of the ODE (with random coefficients)

$$\frac{d}{ds}\tilde{\boldsymbol{\theta}}_s(\mathbf{x}) = \mathbf{G}\big(s,\tilde{\boldsymbol{\theta}}_s(\mathbf{x})\big) = \mathbf{F}\big(s,\tilde{\boldsymbol{\theta}}_s(\mathbf{x})\big) + O\Big(\Big|\int_0^s \sigma(u,\boldsymbol{\chi}_u)dW_u\Big|\Big), \ 0 \le s \le t,$$

with the initial condition $\tilde{\boldsymbol{\theta}}_0(\mathbf{x}) = \mathbf{x}$. By a standard argument of stability, we obtain $\mathbb{E}[|\tilde{\boldsymbol{\chi}}_t - \boldsymbol{\chi}_t|^2] \le C$ and $\mathbb{E}[|\tilde{\boldsymbol{\theta}}_t(\mathbf{x}) - \boldsymbol{\theta}_t(\mathbf{x})|^2] \le C$ for a suitable constant C (depending on (**A**)). Therefore, modifying C if necessary from one line to another, we deduce from (5.14) and (5.16)

$$\begin{aligned} J^q &\ge \mathbb{E}\Big[C^{-1}|\tilde{\boldsymbol{\theta}}_t(\mathbf{x}) - \tilde{\boldsymbol{\chi}}_t|^2 - \ln(q(\boldsymbol{\chi}_t))\Big] \\ &\ge \mathbb{E}\big[C^{-1}|\boldsymbol{\theta}_t(\mathbf{x}) - \boldsymbol{\chi}_t|^2 - \ln(q(\boldsymbol{\chi}_t))\big] - C \\ &\ge \inf_{\mathbf{z}\in\mathbb{R}^{nd}}\big[C^{-1}|\boldsymbol{\theta}_t(\mathbf{x}) - \mathbf{z}|^2 - \ln(q(\mathbf{z}))\big] - C. \end{aligned}$$

Up to a constant depending on a only, we can choose $q(\mathbf{z}) = g_{a,1-t}(\mathbf{z} - \boldsymbol{\theta}_{t,1}(\mathbf{y}))$. One has ($C$ now possibly depending on a)

$$
\begin{aligned}
&J^q \\
&\geq \inf_{\mathbf{z}\in\mathbb{R}^{nd}} \Big[C^{-1}|\boldsymbol{\theta}_t(\mathbf{y}) - \mathbf{z}|^2 + \frac{1-t}{a}|\mathbb{T}_{1-t}^{-1}(\mathbf{z} - \boldsymbol{\theta}_{t,1}(\mathbf{y}))|^2\Big] + \frac{n^2 d}{2}\ln(1-t) - C \\
&\geq C^{-1} \inf_{\mathbf{z}\in\mathbb{R}^{nd}} \Big[|\boldsymbol{\theta}_t(\mathbf{x}) - \mathbf{z}|^2 + (1-t)|\mathbb{T}_{1-t}^{-1}(\mathbf{z} - \boldsymbol{\theta}_{t,1}(\mathbf{y}))|^2\Big] + \frac{n^2 d}{2}\ln(1-t) - C \\
&\geq C^{-1} \inf_{\mathbf{z}\in\mathbb{R}^{nd}} \Big[|\boldsymbol{\theta}_t(\mathbf{x}) - \mathbf{z}|^2 + |\mathbf{z} - \boldsymbol{\theta}_{t,1}(\mathbf{y})|^2\Big] + \frac{n^2 d}{2}\ln(1-t) - C \\
&= (1/2)C^{-1}|\boldsymbol{\theta}_t(\mathbf{x}) - \boldsymbol{\theta}_{t,1}(\mathbf{y})|^2 + \frac{n^2 d}{2}\ln(1-t) - C.
\end{aligned}
$$

Having in mind that $\boldsymbol{\theta}_{t,1}^{-1} = \boldsymbol{\theta}_{1,t}$ is Lipschitz continuous, the Lipschitz constant being uniform in $[0,1)$, we deduce $J^q \geq C^{-1}|\boldsymbol{\theta}_{1,t}(\boldsymbol{\theta}_t(\mathbf{x})) - \mathbf{y}|^2 + (n^2 d/2)\ln(1-t) - C = C^{-1}|\boldsymbol{\theta}_1(\mathbf{x}) - \mathbf{y}|^2 + (n^2 d/2)\ln(1-t) - C$. Finally,

$$
\int_{\mathbb{R}^{nd}} p(t,\mathbf{x},\mathbf{z})(1-t)^{\frac{n^2 d}{2}} g_{a,1-t}(\mathbf{z} - \boldsymbol{\theta}_{t,1}(\mathbf{y}))d\mathbf{z} \leq C\exp(-C|\boldsymbol{\theta}_1(\mathbf{x}) - \mathbf{y}|^2). \qquad \square
$$

A Technical Lemmas

Lemma A.1 *Let $(M_t)_{t\geq 0}$ be a (real-valued) continuous martingale. Then, for any $T > 0$, $\alpha > 0$ and $\beta > 0$, $\mathbb{P}\{\forall t \in [0,T],\ M_t > \alpha + (\beta/2)\langle M\rangle_t\} \leq \exp(-\alpha\beta)$.*

Proof. By Doob's maximal inequality,

$$
\begin{aligned}
&\mathbb{P}\Big\{\forall t \in [0,T],\ M_t > \alpha + (\beta/2)\langle M\rangle_t\Big\} \\
&\quad = \mathbb{P}\Big\{\forall t \in [0,T],\ \exp\Big[\beta M_t - (\beta^2/2)\langle M\rangle_t\Big] > \exp(\alpha\beta)\Big\} \leq \exp(-\alpha\beta).
\end{aligned}
$$

This completes the proof. $\square$

Lemma A.2 *For a real $T > 0$, an integer $i \geq 1$ and a bounded progressively-measurable process $(H_t)_{0\leq t\leq T}$, we set $M_t = \int_0^t (T-s)^{-i} H_s dW_s$, $0 \leq t < T$. Then, for any $\mu \in (0,2]$ and any $\alpha > 0$, with probability greater than $1 - \exp(-\alpha^2)$, for any $t \in [0,T)$,*

$$
(T-t)^i|M_t| \leq (T-t)^{\mu/2}\bigg(\int_0^T \frac{H_s^2 ds}{(T-s)^\mu} + 2\alpha\bigg)^{1/2} \exp\Big(\frac{\alpha}{2}\int_0^T \frac{H_s^2}{(T-s)^\mu} ds\Big).
$$

Proof. By Itô's formula,

$$\begin{aligned} d\Big[(T-t)^{2i-\mu}M_t^2\Big] &= -[2i-\mu](T-t)^{2i-1-\mu}M_t^2dt + (T-t)^{2i-\mu}d[M_t^2] \\ &= -[2i-\mu](T-t)^{2i-1-\mu}M_t^2dt + 2(T-t)^{i-\mu}M_tH_tdW_t \\ &\quad + (T-t)^{-\mu}H_t^2dt. \end{aligned}$$

By Lemma A.1, with probability greater than $1-\exp(-\alpha^2)$,

$$\begin{aligned} (T-t)^{2i-\mu}M_t^2 &\le \int_0^t (T-s)^{-\mu}H_s^2ds + 2\int_0^t (T-s)^{i-\mu}M_sH_sdW_s \\ &\le \int_0^t (T-s)^{-\mu}H_s^2ds + 2\alpha + \alpha\int_0^t (T-s)^{2i-2\mu}M_s^2H_s^2ds. \end{aligned}$$

By Gronwall's lemma, we deduce

$$(T-t)^{2i-\mu}M_t^2 \le \left(\int_0^t (T-s)^{-\mu}H_s^2ds + 2\alpha\right)\exp\left(\alpha\int_0^t \frac{H_s^2}{(T-s)^{\mu}}ds\right). \qquad \square$$

B Auxiliary Gaussian Estimates

Lemma B.1 *With g as in (5.3), for any $a>0$, there exists a constant $c_{B.1}(a)$, depending on a, d and n only, such that for any $\varepsilon,\varepsilon'>0$ and any $\mathbf{x},\mathbf{x}'\in\mathbb{R}^{nd}$,*

$$\begin{aligned} c_{B.1}^{-1}(a)g_{2^{-2n+2}a,\varepsilon+\varepsilon'}(\mathbf{x}-\mathbf{x}') &\le \int_{\mathbb{R}^{nd}} g_{a,\varepsilon}(\mathbf{z}-\mathbf{x})g_{a,\varepsilon'}(\mathbf{z}-\mathbf{x}')d\mathbf{z} \\ &\le c_{B.1}(a)g_{a,\varepsilon+\varepsilon'}(\mathbf{x}-\mathbf{x}'). \end{aligned}$$

Proof (Lemma B.1). Up to a constant C, depending on a, d and n only, the convolution product in the statement is the density at point 0 of the sum of two nd-dimensional independent Gaussian vectors, respectively of mean $\mathbf{x}$ and $-\mathbf{x}'$ and of covariance matrix $[a/2]\varepsilon^{-1}\mathbb{T}_\varepsilon^2$ and $[a/2](\varepsilon')^{-1}\mathbb{T}_{\varepsilon'}^2$. The sum of both vectors has $\mathbf{x}-\mathbf{x}'$ as mean and $[a/2](\varepsilon^{-1}\mathbb{T}_\varepsilon^2+(\varepsilon')^{-1}\mathbb{T}_{\varepsilon'}^2)$ as covariance matrix.

It is well seen that $\varepsilon^{-1}\mathbb{T}_\varepsilon^2+(\varepsilon')^{-1}\mathbb{T}_{\varepsilon'}^2$ is block-diagonal, with n blocks of size $d\times d$. The i^{th} block has the form $[\varepsilon^{2i-1}+(\varepsilon')^{2i-1}]I_d$, where I_d stands for the identity matrix of size d. It is clear that $2^{-2i+2}(\varepsilon+\varepsilon')^{2i-1}\le\varepsilon^{2i-1}+(\varepsilon')^{2i-1}\le(\varepsilon+\varepsilon')^{2i-1}$, where the first bound derives from convexity. Thus, $2^{-2n+2}(\varepsilon+\varepsilon')^{-1}\mathbb{T}_{\varepsilon+\varepsilon'}^2\le\varepsilon^{-1}\mathbb{T}_\varepsilon^2+(\varepsilon')^{-1}\mathbb{T}_{\varepsilon'}^2\le(\varepsilon+\varepsilon')^{-1}\mathbb{T}_{\varepsilon+\varepsilon'}^2$. This proves that $-2^{2n-2}(\varepsilon+\varepsilon')\mathbb{T}_{\varepsilon+\varepsilon'}^{-2}\le-[\varepsilon^{-1}\mathbb{T}_\varepsilon^2+(\varepsilon')^{-1}\mathbb{T}_{\varepsilon'}^2]^{-1}\le-(\varepsilon+\varepsilon')\mathbb{T}_{\varepsilon+\varepsilon'}^{-2}$. Also, $2^{-2n^2d+2nd}(\varepsilon+\varepsilon')^{n^2d}\le\det[\varepsilon^{-1}\mathbb{T}_\varepsilon^2+(\varepsilon')^{-1}\mathbb{T}_{\varepsilon'}^2]\le(\varepsilon+\varepsilon')^{n^2d}$. $\square$

C Another proof of the lower bound via chaining

In this section we provide an alternative proof to derive the lower bound of Theorem 1.1. It uses a chaining argument different from those developed in Kusuoka and Stroock, [KS87], or in the standard uniformly elliptic framework, see e.g [Bas97] Chapter VII. The key idea consists in deriving the lower bound on small balls for the underlying *control metric* (or equivalently on small *ellipsoids* for the Euclidean distance: see Propositions 4.1 and 4.2) by using the parametrix representation (2.17), similarly to [IKO62] in the non-degenerate case. Then, given arbitrary points $\mathbf{x}, \mathbf{x}' \in \mathbb{R}^{nd},\ T > 0$ sufficiently small, we make a suitable chaining of sets so that between two points on consecutive sets we are able to apply the lower bound on small balls.

In the non-degenerate case, the idea was to consider for those chaining sets, balls with centers on the straight line, or geodesic distance, between $\mathbf{x}$ and $\mathbf{x}'$. In our framework we consider (Euclidean) ellipsoids with centers on the optimal path $(\boldsymbol{\phi}_s)_{s\in[0,T]}$ given by Proposition 4.2 (with $t = T$ therein). Precisely, for fixed $\mathbf{x}, \mathbf{x}' \in \mathbb{R}^{nd},\ 0 \le s < t \le T,\ T > 0$ to be specified later on but assumed to be small enough (at least $T \le 1$), Eq. (2.17) yields, for $N \ge 1$,

$$\begin{aligned}
p(s,t,\mathbf{x},\mathbf{x}') \ge \tilde p(s,t,\mathbf{x},\mathbf{x}') - \sum_{k=1}^{N} \int_s^t \int_{\mathbb{R}^{nd}} \tilde p(s,u,\mathbf{x},\mathbf{y}) |H^{\otimes k}(u,t,\mathbf{y},\mathbf{x}')| d\mathbf{y} du \\
- \int_s^t \int_{\mathbb{R}^{nd}} p(s,u,\mathbf{x},\mathbf{y}) |H^{\otimes (N+1)}(u,t,\mathbf{y},\mathbf{x}')| d\mathbf{y} du.
\end{aligned}$$

Now, the second claim in Proposition 5.1 still holds with $(1,t)$ replaced by (t,u), so that

$$\begin{aligned}
&p(s,t,\mathbf{x},\mathbf{x}') \ge \tilde p(s,t,\mathbf{x},\mathbf{x}') \\
&- \sum_{k=1}^{N} C(k) \int_s^t \int_{\mathbb{R}^{nd}} \tilde p(s,u,\mathbf{x},\mathbf{y}) (t-u)^{k\frac{\eta}{2}-1} g_{C(k),t-u}(\mathbf{y} - \boldsymbol{\theta}_{u,t}(\mathbf{x}')) d\mathbf{y} du \\
&- C(N+1) \int_s^t \int_{\mathbb{R}^{nd}} p(s,u,\mathbf{x},\mathbf{y}) (t-u)^{(N+1)\frac{\eta}{2}-1} g_{C(N+1),t-u}(\mathbf{y} - \boldsymbol{\theta}_{u,t}(\mathbf{x}')) d\mathbf{y} du.
\end{aligned}$$

Taking $\lfloor N\eta/2 \rfloor \ge n^2 d/2 + 1$, Proposition 5.2 (replacing $(1,t)$ by (t,u) as well), Lemma (B.1) and a suitable version of (2.13) give

$$\begin{aligned}
p(s,t,\mathbf{x},\mathbf{x}') &\ge \tilde p(s,t,\mathbf{x},\mathbf{x}') \\
&\quad - C\frac{(t-s)^{\eta/2}}{(t-s)^{n^2 d/2}} \exp(-C^{-1}(t-s)|\mathbb{T}_{t-s}^{-1}(\boldsymbol{\theta}_{t,s}(\mathbf{x}) - \mathbf{x}')|^2) \\
&\ge \frac{C^{-1}}{(t-s)^{n^2 d/2}} \exp(-C(t-s)|\mathbb{T}_{t-s}^{-1}(\boldsymbol{\theta}_{t,s}(\mathbf{x}) - \mathbf{x}')|^2) \\
&\quad - C\frac{(t-s)^{\eta/2}}{(t-s)^{n^2 d/2}} \exp(-C^{-1}(t-s)|\mathbb{T}_{t-s}^{-1}(\boldsymbol{\theta}_{t,s}(\mathbf{x}) - \mathbf{x}')|^2),
\end{aligned}$$

for $C := C(\mathbf{A})$ large enough, where the lower bound for $\tilde{p}$ can be derived similarly to the upper bound given in Lemma 5.4. Set now for simplicity $d_{s,t}(\mathbf{x},\mathbf{x}') := (t-s)^{1/2}|\mathbb{T}_{t-s}^{-1}(\boldsymbol{\theta}_{t,s}(\mathbf{x})-\mathbf{x}')|$. If $d_{s,t}(\mathbf{x},\mathbf{x}') \le 1$, then $p(s,t,\mathbf{x},\mathbf{x}') \ge (t-s)^{-n^2d/2}[C^{-1}\exp(-C) - CT^{\eta/2}]$. Hence, for $T \le (2C^2)^{-2/\eta}\exp(-2C/\eta)$,

$$p(s,t,\mathbf{x},\mathbf{x}') \ge C_0(t-s)^{-n^2d/2},\ C_0 = \exp(-C)(2C)^{-1}. \tag{C.1}$$

Equation (C.1) is what we call the lower bound on balls w.r.t. to the control metric, for T small enough.

Now, for $\mathbf{x},\mathbf{x}' \in \mathbb{R}^{nd}$, if $d_T := d_{0,T}(\mathbf{x},\mathbf{x}') \le 1$, then the previous bound applies. If $d_T \ge 1$, we need to do a chaining. We are going to define sets whose centers are equally distributed w.r.t. to the level sets of the energy associated to the optimal path $(\boldsymbol{\phi}_s)_{s\in[0,T]}$, $\boldsymbol{\phi}_0 = \mathbf{x}$, $\boldsymbol{\phi}_T = \mathbf{x}'$, constructed in Proposition 4.2 (with $t = T$ therein). Recalling that $(\varphi_s)_{s\in[0,T]}$ denotes the associated optimal control, we define $t_0 = 0$ and, for all $i \ge 1$:

$$t_i \begin{cases} := \inf\Big\{t \in [t_{i-1},T] : \displaystyle\int_{t_{i-1}}^t |\varphi_s|^2 ds = \frac{I(T,\mathbf{x},\mathbf{x}')}{M_0}\Big\} \wedge (t_{i-1} + \frac{T}{M_0}) \\ \qquad\qquad \text{if } t_{i-1} < T(1-\dfrac{2}{M_0}) \\ := T \quad \text{if } t_{i-1} \ge T(1-\dfrac{2}{M_0}), \end{cases}$$

where $M_0 := \lceil K d_T^2 \rceil \ge 3$ for $K \ge 3$ to be specified later on. Set now, for all $i \ge 1$, $\varepsilon_i := t_{i+1} - t_i$. We have

Lemma C.1 (Controls on the time step) *There exist a constant $C_1 := C_1(\mathbf{A}) \le 1$ and an integer $M_1 \in [M_0/2, M_0/C_1]$, s.t. $t_{M_1} = T$ and*

$$\forall i \in \{0,\cdots,M_1-1\},\ C_1\frac{T}{M_0} \le \varepsilon_i \le 2\frac{T}{M_0}. \tag{C.2}$$

Proof. We first set $M_1 = \inf\{k \ge 1 : t_k = T\}$. (The set $\{k \ge 1 : t_k = T\}$ is clearly non-empty.) The upper bound in (C.2) then follows from the definition of the family $(t_i)_{i\ge1}$. Suppose now that $t_i < T(1-2/M_0)$ for a given $0 \le i \le M_1 - 1$. Assume also that $t_{i+1} - t_i < T/M_0$ (otherwise $\varepsilon_i = T/M_0$). Then, $\int_{t_i}^{t_{i+1}} |\varphi_s|^2 ds = I(T,\mathbf{x},\mathbf{x}')/M_0$. From Proposition 4.2, (1), we recall that

$$\sup_{0\le s\le T} |\varphi_s| \le C_2|\mathbb{T}_T^{-1}[\boldsymbol{\theta}_T(\mathbf{x}) - \mathbf{x}']| = C_2 T^{-1/2} d_T,$$

where $C_2 = C_2(\mathbf{A})$. Hence,

$$\frac{I(T,\mathbf{x},\mathbf{x}')}{M_0} = \int_{t_i}^{t_{i+1}} |\varphi_s|^2 ds \le C_2 \varepsilon_i T^{-1} d_T^2.$$

Modifying C_2 if necessary, Propositions 4.1 and 4.2 yield $d_T^2/M_0 \le C_2\varepsilon_i T^{-1} d_T^2$ and the lower bound in (C.2) follows for all i s.t. $t_i < T(1-2/M_0)$. The bound for M_1 is then easily derived. $\square$

Introduce now for all $i \in \{0, \cdots, M_1\}$, $\mathbf{y}_i = \boldsymbol{\phi}_{t_i}$ (in particular $\mathbf{y}_0 = \mathbf{x}, \mathbf{y}_{M_1} = \mathbf{x}'$), and for all $i \in \{1, \cdots, M_1 - 1\}$,

$$B_i := \Big\{\mathbf{z} \in \mathbb{R}^{nd} : \\ K^{1/2}\rho\Big(|\mathbb{T}_{K\rho^2}^{-1}(\boldsymbol{\theta}_{t_i,t_{i-1}}(\mathbf{y}_{i-1}) - \mathbf{z})| + |\mathbb{T}_{K\rho^2}^{-1}(\mathbf{z} - \boldsymbol{\theta}_{t_i,t_{i+1}}(\mathbf{y}_{i+1}))| \le K^{-1/2}\Big\},$$

where $\rho := T^{1/2}d_T/M_0$. Write now (with $\mathbf{x}_0 = \mathbf{x}$ and $\mathbf{x}_{M_1} = \mathbf{x}'$)

$$p(T, \mathbf{x}, \mathbf{x}') \ge \int_{\prod_{i=1}^{M_1-1} B_i} \prod_{i=0}^{M_1-1} p(t_i, t_{i+1}, \mathbf{x}_i, \mathbf{x}_{i+1}) d\mathbf{x}_1 d\mathbf{x}_2 \cdots d\mathbf{x}_{M_1-1}. \tag{C.3}$$

The following lemma, whose proof is postponed to the end of the section allows to derive the lower bound.

Lemma C.2 (Controls for the chaining) *With the previous assumptions and definitions, there exists a constant K_0, depending on the parameters in (**A**) only, such that, for $K \ge K_0$,*

$$\begin{aligned} &\forall i \in \{1, \cdots, M_1 - 2\},\ \forall (\mathbf{x}_i, \mathbf{x}_{i+1}) \in B_i \times B_{i+1}, \\ &\qquad\qquad\qquad \varepsilon_i^{1/2}|\mathbb{T}_{\varepsilon_i}^{-1}(\boldsymbol{\theta}_{t_{i+1},t_i}(\mathbf{x}_i) - \mathbf{x}_{i+1})| \le 1, \\ &\forall \mathbf{x}_1 \in B_1, \qquad\quad \varepsilon_0^{1/2}|\mathbb{T}_{\varepsilon_0}^{-1}(\boldsymbol{\theta}_{t_1,0}(\mathbf{x}) - \mathbf{x}_1)| \le 1, \\ &\forall \mathbf{x}_{M_1-1} \in B_{M_1-1}, \quad \varepsilon_{M_1-1}^{1/2}|\mathbb{T}_{\varepsilon_{M_1-1}}^{-1}(\boldsymbol{\theta}_{T,t_{M_1-1}}(\mathbf{x}_{M_1-1}) - \mathbf{x}')| \le 1. \end{aligned} \tag{C.4}$$

Moreover, for $K \ge K_0$ and for the same C_1 as in Lemma C.1,

$$\forall i \in \{1, \cdots, M_1 - 1\}, \quad |B_i| \ge C_1 \rho^{n^2 d}, \tag{C.5}$$

where $|B_i|$ stands for the Lebesgue measure of the set B_i.

We assume that $K \ge K_0$. We deduce from Eqs. (C.1), (C.3) and (C.4) that

$$p(T, \mathbf{x}, \mathbf{x}') \ge \frac{C_0}{\varepsilon_0^{n^2d/2}} \prod_{i=1}^{M_1-1} \frac{C_0}{\varepsilon_i^{n^2d/2}} |B_i|. \tag{C.6}$$

Recalling $\rho = T^{1/2}d_T/M_0$ and plugging (C.5), (C.2) into (C.6) we derive

$$p(T, \mathbf{x}, \mathbf{x}') \ge \Big(\frac{C_0 C_1}{2^{n^2d/2}}\Big)^{M_1} \Big(\frac{M_0}{T}\Big)^{n^2d/2} \left(\frac{d_T^2}{M_0}\right)^{(M_1-1)(n^2d/2)}.$$

Now, by definition of M_0, $M_0 - 1 \le K d_T^2$ so that $d_T^2/M_0 \ge 1/(K+1)$ since $d_T \ge 1$. Setting $C_3 := (C_0C_1)/(2(K+1))^{n^2d/2} < 1$ for K large enough, we obtain

$$
\begin{aligned}
p(T,\mathbf{x},\mathbf{x}') &\geq C_3^{M_1}(K+1)^{M_1 n^2 d/2}\Big(\frac{M_0}{T}\Big)^{n^2 d/2}(K+1)^{-(M_1-1)(n^2 d/2)}\\
&= T^{-n^2 d/2}(K+1)^{n^2 d/2} M_0^{n^2 d/2} C_3^{M_1}\\
&\geq T^{-n^2 d/2} C_3^{M_1}\\
&= T^{-n^2 d/2}\exp(-\ln(1/C_3)M_1).
\end{aligned}
$$

By Lemma C.1, we know that $M_1 \leq M_0/C_1$, so that

$$
\begin{aligned}
p(T,\mathbf{x},\mathbf{x}') &\geq T^{-n^2 d/2}\exp\Big(-[\ln(1/C_3)/C_1]M_0\Big)\\
&= T^{-n^2 d/2}\exp\Big(-[\ln(1/C_3)/C_1]\Big)\exp\Big(-[\ln(1/C_3)/C_1](M_0-1)\Big)\\
&\geq T^{-n^2 d/2}\exp\Big(-[\ln(1/C_3)/C_1]\Big)\exp\Big(-[\ln(1/C_3)/C_1]K d_T^2\Big).
\end{aligned}
$$

Eventually, we can choose K greater than K_0 such that C_3 above be strictly less than one. For a such a choice, the parameters in the above lower bound depends on $(\mathbf{A})$ only as required. This proves the lower bound in short time (see (C.1)). By a scaling argument, we obtain the lower bound on any arbitrary interval. $\square$

Proof of Lemma C.2. Let us begin with the proof of (C.4). Fix $(\mathbf{x}_i,\mathbf{x}_{i+1}) \in B_i \times B_{i+1},\ i \in \{1,\cdots,M_1-2\}$. By (2.13), we can find $C_4 = C_4(\mathbf{A})$ such that

$$
\begin{aligned}
Q_i &:= \varepsilon_i^{1/2}|\mathbb{T}_{\varepsilon_i}^{-1}(\boldsymbol{\theta}_{t_{i+1},t_i}(\mathbf{x}_i)-\mathbf{x}_{i+1})|\\
&\leq C_4\varepsilon_i^{1/2}|\mathbb{T}_{\varepsilon_i}^{-1}(\mathbf{x}_i-\boldsymbol{\theta}_{t_i,t_{i+1}}(\mathbf{x}_{i+1}))|\\
&\leq C_4\varepsilon_i^{1/2}\Big\{|\mathbb{T}_{\varepsilon_i}^{-1}(\mathbf{x}_i-\boldsymbol{\theta}_{t_i,t_{i-1}}(\mathbf{y}_{i-1}))| + |\mathbb{T}_{\varepsilon_i}^{-1}(\boldsymbol{\theta}_{t_i,t_{i-1}}(\mathbf{y}_{i-1})-\mathbf{y}_i)|\\
&\quad + |\mathbb{T}_{\varepsilon_i}^{-1}(\mathbf{y}_i-\boldsymbol{\theta}_{t_i,t_{i+1}}(\mathbf{x}_{i+1}))|\Big\}\\
&:= Q_i^1+Q_i^2+Q_i^3.
\end{aligned}
$$

One has

$$
\begin{aligned}
Q_i^1 &\leq C_4\sum_{j=1}^n \varepsilon_i^{1/2-j}|(\mathbf{x}_i-\boldsymbol{\theta}_{t_i,t_{i-1}}(\mathbf{y}_{i-1}))_j|\\
&\leq C_4\sum_{j=1}^n\left(\frac{\varepsilon_i}{K\rho^2}\right)^{1/2-j}(K\rho^2)^{1/2-j}|(\mathbf{x}_i-\boldsymbol{\theta}_{t_i,t_{i-1}}(\mathbf{y}_{i-1}))_j|.
\end{aligned}
\tag{C.7}
$$

By (C.2), $\varepsilon_i/(K\rho^2) \geq C_1(T/M_0)/(KTd_T^2/M_0^2) = C_1M_0/(Kd_T^2) \geq C_1$. Thus, for all $j\in\{1,\cdots,n\}$, $(\varepsilon_i/(K\rho^2))^{1/2-j} \leq C_1^{1/2-j} \leq C_1^{-n}$ and

$$
\begin{aligned}
Q_i^1 &\leq C_4C_1^{-n}\sum_{j=1}^n(K\rho^2)^{1/2-j}|(\mathbf{x}_i-\boldsymbol{\theta}_{t_i,t_{i-1}}(\mathbf{y}_{i-1}))_j|\\
&= C_4C_1^{-n}K^{1/2}\rho|\mathbb{T}_{K\rho^2}^{-1}(\mathbf{x}_i-\boldsymbol{\theta}_{t_i,t_{i-1}}(\mathbf{y}_{i-1}))| \leq C_4C_1^{-n}K^{-1/2},
\end{aligned}
$$

exploiting $\mathbf{x}_i \in B_i$ for the last identity. The term Q_i^3 could be handled in a similar way so that $Q_i^1 + Q_i^3 \le C_4 C_1^{-n} K^{-1/2}$. Now, by Proposition 4.1, there exists a constant $C_5 := C_5(\mathbf{A})$ such that

$$
\begin{aligned}
Q_i^2 &\le C_5 I(t_i, t_{i+1}, \mathbf{y}_i, \mathbf{y}_{i+1}) \\
&\le C_5 \int_{t_i}^{t_{i+1}} |\varphi_s|^2 ds \le C_5 \frac{I(T, \mathbf{x}, \mathbf{x}')}{M_0} \le C_5^2 \frac{d_T^2}{M_0} \le \frac{C_5^2}{K}.
\end{aligned} \tag{C.8}
$$

Hence, for all $i \in \{1, \cdots, M_1 - 2\}$, $Q_i \le 1$ for K large enough (the expression "large enough" refers to the parameters in ($\mathbf{A}$) only). Eventually, for $\mathbf{x}_1 \in B_1, \mathbf{x}_{M_1-1} \in B_{M_1-1}$ the terms $Q_0 := \varepsilon_0^{1/2} |\mathbb{T}_{\varepsilon_0}^{-1}(\boldsymbol{\theta}_{t_1,0}(\mathbf{x}) - \mathbf{x}_1)|$ and $Q_{M_1-1} := \varepsilon_{M_1-1}^{1/2} |\mathbb{T}_{\varepsilon_{M_1-1}}^{-1}(\boldsymbol{\theta}_{T, t_{M_1-1}}(\mathbf{x}_{M_1-1}) - \mathbf{x}')| \le C_4 \varepsilon_{M_1-1}^{1/2} |\mathbb{T}_{\varepsilon_{M_1-1}}^{-1}(\mathbf{x}_{M_1-1} - \boldsymbol{\theta}_{t_{M_1-1}, T}(\mathbf{x}'))|$ can be controlled as the previous $Q_i^1,\ i \in \{1, \cdots, M_1 - 2\}$ from the definitions of B_1, B_{M_1-1}, so that $Q_i \le 1,\ i \in \{0, M_1 - 1\}$ as well. This proves (C.4).

It now remains to control the Lebesgue measure of the sets $(B_i)_{i \in \{1, \cdots, M_1-1\}}$. Define for all $i \in \{1, \cdots, M_1 - 1\}$, $E_i := \{\mathbf{z} \in \mathbb{R}^{nd} : K^{1/2} \rho |\mathbb{T}_{K\rho^2}^{-1}(\mathbf{y}_i - \mathbf{z})| \le (1/3) K^{-1/2}\}$. Then, $|E_i| = C_6 3^{-nd} K^{-nd/2} (K\rho^2)^{n^2 d/2}$ for a universal constant C_6 depending on n and d only. Modifying C_6 if necessary, we obtain $|E_i| \ge C_6 \rho^{n^2 d}$ for $K \ge 1$. Let us now prove $E_i \subset B_i$. Write, for all $\mathbf{z} \in E_i$,

$$
\begin{aligned}
R_i &:= K^{1/2} \rho \Big(|\mathbb{T}_{K\rho^2}^{-1}(\boldsymbol{\theta}_{t_i, t_{i-1}}(\mathbf{y}_{i-1}) - \mathbf{z})| + |\mathbb{T}_{K\rho^2}^{-1}(\mathbf{z} - \boldsymbol{\theta}_{t_i, t_{i+1}}(\mathbf{y}_{i+1}))| \Big) \\
&\le K^{1/2} \rho \Big(|\mathbb{T}_{K\rho^2}^{-1}(\boldsymbol{\theta}_{t_i, t_{i-1}}(\mathbf{y}_{i-1}) - \mathbf{y}_i)| + 2 |\mathbb{T}_{K\rho^2}^{-1}(\mathbf{y}_i - \mathbf{z})| \\
&\quad + |\mathbb{T}_{K\rho^2}^{-1}(\mathbf{y}_i - \boldsymbol{\theta}_{t_i, t_{i+1}}(\mathbf{y}_{i+1}))| \Big) \\
&:= R_i^1 + R_i^2 + R_i^3.
\end{aligned}
$$

By definition of E_i, $R_i^2 \le (2/3) K^{-1/2}$. Following (C.7) and (C.8), there exists a constant $C_7 := C_7(\mathbf{A}) > 0$ such that

$$
\begin{aligned}
R_i^1 &\le C_7 \sum_{j=1}^{n} \left(\frac{\varepsilon_i}{K\rho^2} \right)^{j-1/2} \varepsilon_i^{1/2-j} |(\boldsymbol{\theta}_{t_i, t_{i-1}}(\mathbf{y}_{i-1}) - \mathbf{y}_i)_j| \\
&\le C_7 \Big(\frac{2M_0}{K d_T^2} \Big)^n \varepsilon_i^{1/2} |\mathbb{T}_{\varepsilon_i}^{-1}(\boldsymbol{\theta}_{t_i, t_{i-1}}(\mathbf{y}_{i-1}) - \mathbf{y}_i)| \\
&\le C_7^2 \Big(\frac{2(K d_T^2 + 1)}{K d_T^2} \Big)^n K^{-1} \le C_7^2 4^n K^{-1}.
\end{aligned}
$$

Since the term R_i^3 can be handled in the same way, we deduce that for K large enough $R_i \le K^{-1/2}$, so that $E_i \subset B_i$. This completes the proof. □

Acknowledgment

We would like to thank Valentin Konakov for fruitful discussion during the preparation of this work.

References

[Aro67] D. G. Aronson. Bounds for the fundamental solution of a parabolic equation. *Bull. Amer. Math. Soc.*, 73:890–896, 1967.

[Bal90] V. Bally. On the connection between the Malliavin covariance matrix and Hörmander's condition. *J. Funct. Anal.*, 96–2:219–255, 1990.

[Bas97] R. F. Bass. *Diffusions and Elliptic Operators.* Springer, 1997.

[BKH09] V. Bally and A. Kohatsu-Higa. Lower bounds for densities of Asian typestochastic differential equations. *to appear in J. Funct. Anal.*, 2009.

[BL91] G. Ben Arous and R. Léandre. Décroissance exponentielle du noyau de la chaleur sur la diagonale. II. *Prob. Th. Rel. Fields*, 90(3):377–402, 1991.

[BL08] T. Bodineau and L. Lefevere. Large deviations of lattice Hamiltonian dynamics coupled to stochastic thermostats. *Journal of Statistical Physics*, 133:1–27, 2008.

[BP07] U. Boscain and S. Polidoro. Gaussian estimates for hypoelliptic operators via optimal control. *Atti Accad. Naz. Lincei*, 18–4:333–342, 2007.

[BPV01] E. Barucci, S. Polidoro, and V. Vespri. Some results on partial differential equations and asian options. *Math. Models Methods Appl. Sci*, 3:475–497, 2001.

[Cat90] P. Cattiaux. Calcul stochastique et opérateurs dégénérés du second ordre - I. Résolvantes, théorème de Hörmander et applications. *Bull. Sc. Math., 2ème série*, 114:421–462, 1990.

[CM02] P. Cattiaux and L. Mesnager. Hypoelliptic non-homogeneous diffusions. *Prob. Th. Rel. Fields*, 123(4):453–483, 2002.

[Cor07] J.-M. Coron. *Control and nonlinearity.* Mathematical Surveys and Monographs, 136, AMS, 2007.

[EPRB99] J.-P. Eckmann, C.-A. Pillet, and L. Rey-Bellet. Non-equilibrium statistical mechanics of anharmonic chains coupled to two heat baths at different temperatures. *Comm. Math. Phys.*, 201–3:657–697, 1999.

[FP05] M. Di Francesco and A. Pascucci. On a class of degenerate parabolic equations of kolmogorov type. *Applied Mathematics Research eXpress*, 3:77–116, 2005.

[FP06] M. Di Francesco and S. Polidoro. Schauder estimates, Harnack inequality and Gaussian lower bound for kolmogorov-type operators in non-divergence form. *Adv. Differential Equations*, 11:1261–1320, 2006.

[Fri64] A. Friedman. *Partial differential equations of parabolic type.* Prentice-Hall, 1964.

[FW98] M. Freidlin and A. Wentzell. *Random Perturbations of Dynamical Systems.* Springer, 1998.

[HN04] F. Hérau and F. Nier. Isotropic hypoellipticity and trend to equilibrium for the Fokker-Planck equation with a high-degree potential. *Arch. Ration. Mech. Anal.*, 171–2:151–218, 2004.

[Hör67] L. Hörmander. Hypoelliptic second order differential operators. *Acta. Math.*, 119:147–171, 1967.

[IKO62] A. M. Il'in, A. S. Kalashnikov, and O. A. Oleinik. Second-order linear equations of parabolic type. *Uspehi Mat. Nauk*, 17–3(105):3–146, 1962.

[IL90] H. Ishii and P.-L. Lions. Viscosity solutions of fully nonlinear second-order elliptic partial differential equations. *J. Diff. Equations*, 83:26–78, 1990.

[KM00] V. Konakov and E. Mammen. Local limit theorems for transition densities of Markov chains converging to diffusions. *Prob. Th. Rel. Fields*, 117:551–587, 2000.

[KMM09] V. Konakov, S. Menozzi, and S. Molchanov. Explicit parametrix and local limit theorems for some degenerate diffusion processes. *To appear in Annales de l'Institut Henri Poincaré, Série B*, 2009.

[Kol34] A. N. Kolmogorov. Zufällige Bewegungen (zur Theorie der Brownschen Bewegung). *Ann. of Math.*, 2-35:116–117, 1934.

[KS84] S. Kusuoka and D. Stroock. Applications of the Malliavin calculus. I. *Stochastic analysis (Katata/Kyoto, 1982)*, North-Holland Math. Library, 32:271–306, 1984.

[KS85] S. Kusuoka and D. Stroock. Applications of the Malliavin calculus. II. *J. Fac. Sci. Univ. Tokyo Sect. IA Math*, 32:1–76, 1985.

[KS87] S. Kusuoka and D. Stroock. Applications of the Malliavin calculus. III. *J. Fac. Sci. Univ. Tokyo Sect. IA Math*, 34:391–442, 1987.

[Mal78a] P. Malliavin. C^k-hypoellipticity with degeneracy. *In: Stochastic analysis (Proc. Internat. Conf., Northwestern Univ., Evanston, Ill., 1978), Academic Press, New York-London*, pages 199–214, 1978.

[Mal78b] P. Malliavin. Stochastic calculus of variation and hypoelliptic operators. *In: Proceedings of the International Symposium on Stochastic Differential Equations (Res. Inst. Math. Sci., Kyoto Univ., Kyoto, 1976), Wiley, New York-Chichester-Brisbane*, pages 195–263, 1978.

[Mal97] P. Malliavin. *Stochastic analysis.* Springer-Verlag, Berlin, 1997.

[MS67] H. P. McKean and I. M. Singer. Curvature and the eigenvalues of the Laplacian. *J. Differential Geometry*, 1:43–69, 1967.

[MZ08] J. Ma and J. Zhang. Weak solutions for forward-backward SDEs–a martingale problem approach. *Ann. Probab.*, 36:2092–2125, 2008.

[Nua95] D. Nualart. *The Malliavin calculus and related topics. Probability and its Applications.* Springer-Verlag, New York, 1995.

[Pol94] S. Polidoro. On a class of ultraparabolic operators of Kolmogorov-Fokker-Planck type. *Le Matematiche*, 49:53–105, 1994.

[PP06] A. Pascucci and S. Polidoro. Harnack inequalities and Gaussian estimates for a class of hypoelliptic operators. *Trans. Amer. Math. Soc.*, 358–11:4873–4893, 2006.

[RBL00] L. Rey-Bellet and L.E. Lawrence. Asymptotic behavior of thermal nonequilibrium steady states for a driven chain of anharmonic oscillators. *Comm. Math. Phys.*, 215–1:1–24, 2000.

[RY99] D. Revuz and M. Yor. *Continuous martingales and Brownian motion. 3rd ed.* Grundlehren der Mathematischen Wissenschaften. 293. Berlin: Springer, 1999.

[She91] S. J. Sheu. Some estimates of the transition density of a nondegenerate diffusion Markov process. *Ann. Probab.*, 19–2:538–561, 1991.

[Soi94] C. Soize. *The Fokker-Planck equation for stochastic dynamical systems and its explicit steady state solutions.* Series on Advances in Mathematics for Applied Sciences, 17. World Scientific Publishing Co., Inc., River Edge, NJ, 1994.

[Son67] I. M. Sonin. A class of degenerate diffusion processes (russian). *Teor. Verojatnost. i Primenen*, 12:540–547, 1967.

[Son98] E. D. Sontag. *Mathematical control theory, second ed.* Springer, 1998.

[Str83] D.W. Stroock. Some applications of stochastic calculus to partial differential equations. *In: Eleventh Saint Flour probability summer school—1981 (Saint Flour, 1981), Lecture Notes in Math., 976, Springer, Berlin*, pages 267–382, 1983.

[SV79] D.W. Stroock and S.R.S. Varadhan. *Multidimensional diffusion processes.* Springer-Verlag Berlin Heidelberg New-York, 1979.

[Tal02] D. Talay. Stochastic Hamiltonian dissipative systems: exponential convergence to the invariant measure, and discretization by the implicit Euler scheme. *Markov Processes and Related Fields*, 8–2:163–198, 2002.

[Web51] M. Weber. The fundamental solution of a degenerate partial differential equation of parabolic type. *Trans. Amer. Math. Soc.*, 71:24–37, 1951.